

Access DB#

85268

**SEARCH REQUEST FORM**

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
 Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

**If more than one search is submitted, please prioritize searches in order of need.**

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Jan Delaval  
 Reference Librarian  
 Biotechnology & Chemical Library  
 CM1 1E07 - 703-308-4498  
 jan.delaval@uspto.gov

**STAFF USE ONLY****Type of Search****Vendors and cost where applicable**

Searcher: Ch  
 Searcher Phone #: 4458  
 Searcher Location: \_\_\_\_\_  
 Date Searcher Picked Up: 1/27/03  
 Date Completed: 2/2/03  
 Searcher Prep & Review Time: \_\_\_\_\_  
 Clerical Prep Time: 30  
 Online Time: 75

NA Sequence (#) <input checked="" type="checkbox"/>	STN _____
AA Sequence (#) <input checked="" type="checkbox"/>	Dialog _____
Structure (#) _____	Questel/Orbit _____
Bibliographic _____	Dr.Link _____
Litigation _____	Lexis/Nexis _____
Fulltext _____	Sequence Systems <input checked="" type="checkbox"/>
Patent Family _____	WWW/Internet _____
Other _____	Other (specify) _____





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:23 ; Search time 26.5 Seconds  
(without alignments)  
1749.457 Million cell updates/sec

Title: US-09-931-704-2  
Perfect score: 1226  
Sequence: 1 MDLRAGDSWGLACLCTVLW.....KKMQPPAAAVTLHLGAHGF 225

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1226	100.0	225	4	Q9UBD9
2	1193	97.3	225	11	Q9QZM3
3	150.5	12.3	215	13	Q9PUJ2
4	150.5	12.3	215	13	Q9PUJ1
5	150.5	12.3	215	13	Q9PUJ0
6	148.5	12.1	215	13	Q9PUJ9
7	97.5	8.0	318	4	Q96LS2
8	97.5	8.0	530	3	Q8X0E9
9	96.5	7.9	455	11	Q9CWV7
10	92	7.5	332	10	Q9MAV1
11	92	7.5	423	11	Q9JHE4
12	90	7.3	423	11	Q9D8V6
13	89.5	7.3	771	2	Q9S3Q9
14	89	7.3	733	16	Q9I664
15	87	7.1	955	11	Q88287
16	87	7.1	1561	11	O88286

17	86.5	7.1	640	16	O54153	streptomyc
18	85.5	7.0	1154	11	Q921R2	mus musculu
19	85	6.9	476	11	Q8R363	mus musculu
20	85	6.9	727	11	O88841	mus musculu
21	84.5	6.9	294	11	Q9CP21	mus musculu
22	84.5	6.9	389	17	Q9YA71	aeropytrum p
23	84	6.9	200	16	O8YB81	brucella me
24	84	6.9	294	4	Q9NX5	homo sapien
25	84	6.9	411	16	Q9RVF3	deinococcus
26	84	6.9	559	16	O83432	treponema p
27	84	6.9	876	4	Q9P210	homo sapien
28	83.5	6.8	310	4	O9S159	homo sapien
29	83.5	6.8	5990	2	Q9RLP6	mycobacteri
30	83	6.8	222	2	Q9L584	streptococc
31	82.5	6.7	310	4	Q9UID0	homo sapien
32	82.5	6.7	542	16	Q9HTB5	pseudomonas
33	82	6.7	283	2	Q9ZB87	pseudomonas
34	82	6.7	611	16	Q98LP5	rhizobium l
35	82	6.7	721	4	Q96PC8	homo sapien
36	82	6.7	756	4	Q96PC7	homo sapien
37	82	6.7	1179	10	O65507	arabidopsis
38	82	6.7	2376	10	Q9FIN7	arabidopsis
39	81.5	6.6	396	2	Q9AF00	arabidopsis
40	81	6.6	328	15	Q9EI25	human immun
41	81	6.6	328	15	Q9EI22	human immun
42	81	6.6	479	16	O8UG24	rhodotermu
43	81	6.6	924	2	Q9X582	rhodotermu
44	80.5	6.6	305	4	Q9H7R6	homo sapien
45	80.5	6.6	353	16	Q8Z0G6	anabaena sp

## ALIGNMENTS

## RESULT 1

ID	Q9UBD9	PRELIMINARY;	PRT;	225 AA.
AC	Q9UBD9;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	Neurotrophin-1/B-cell stimulating factor-3 (Cardiotrophin-like cytokine)			
DE	(Similar to cardiotrophin-like cytokine, neurotrophin-1/B-cell stimulating factor-3).			
GN	CLC.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NCBI_TaxID=9606;				
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=99432254; PubMed=10500198;			
RA	Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S.,			
RA	Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,			
RA	Simonet W.S., Boone T., Chang M.-S.;			
RT	"Novel neurotrophin-1/B cell-stimulating factor-3: A cytokine of the IL-6 family."			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).			
RN	[2]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=99382254; PubMed=10448081;			
RX	Shi Y., Wang W., Yourey P.A., Gohari S., Zukauskas D., Zhang J.,			
RA	Ruben S., Alderson R.F.;			
RT	"Computational EST database analysis identifies a novel member of the neurotrophic cytokine family."			
RL	Biochem. Biophys. Res. Commun. 262:132-138(1999).			
RN	[3]			
RN	SEQUENCE FROM N.A.			
RA	Hu X., Xu Y., Zhang B., Peng X., Yuan J., Qiang B.;			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RP	TISSUE=KIDNEY;			





Db 235 LPSRSTPSPPTVES 249

RESULT 8

Q8XOE9 PRELIMINARY; PRT; 530 AA.

AC Q8XOE9; 7.9%; Score 96.5; DB 11; Length 455;

DT 01-MAR-2002 (TREMELrel. 20, Created)

DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)

DE Hypothetical 59.4 kDa protein.

GN B14A6.080.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariales; Sordariaceae; Neurospora.

OX NCBI\_TaxID=5141;

EN [1];

RP SEQUENCE FROM N.A.

RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.W., Mannhaupt G.,

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

RN [2];

RP SEQUENCE FROM N.A.

RA German Neurospora genome project;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

RL EMBL; AL670007; CAD21310.1; -.

DR InterPro; IPR001810; F-box.

DR PROSITE; PS0181; FBOX; 1.

KW Hypothetical protein.

SQ SEQUENCE 530 AA; 59352 MW; 70382EEL5F71BB9D CRC64;

Query Match 8.0%; Score 97.5; DB 3; Length 530;

Best Local Similarity 24.8%; Pred. No. 1;

Matches 41; Conservative 22; Mismatches 65; Indels 37; Gaps 6;

QY 64 YLGGPPNEPDPNPRCLGAETLPA-TVDLEWRSINDKRLRNTQYAYSHLLCYLRLGNL 122

Db 236 FLVPPVLEPDPFLVLAGLKKLHLVDLADLRDMEIQYLLFKKFLALTPNLTLWL 291

QY 123 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGVPLPQLPCT-----EPTWTPGP 173

Db 292 ----VNPERISISHY-----GGKEALLKWLASSLRGTWSSPGLINADPSRLPPP 336

QY 174 AHSDFLQKMDDFWLLKELQTLWPSAKDFNRLKKMKOPPAATAVL 218

Db 337 VEFHLEQLD-----IGQLDV-----SANTLYLRFNKSSTLKAISL 373

RESULT 9

Q9CWV7 PRELIMINARY; PRT; 455 AA.

AC Q9CWV7; 7.9%; Score 96.5; DB 11; Length 455;

DT 01-JUN-2001 (TREMELrel. 17, Created)

DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)

DE 2410003H12RIK protein.

GN 2410003H12RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

EN [1];

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,

RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Glissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,

Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,

Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

Ra Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,

Hayashizaki Y.,

RT "Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690 (2001).

DR EMBL; AK010358; BAB26878.1; -.

DR MGI; MGI:1919221; 2410003H12RIK.

SQ SEQUENCE 455 AA; 51481 MW; 201886B814EB9CFF CRC64;

Query Match 7.9%; Score 96.5; DB 11; Length 455;

Best Local Similarity 22.4%; Pred. No. 1.1;

Matches 59; Conservative 24; Mismatches 89; Indels 91; Gaps 12;

QY 7 DSWGMLACLCTVLW-HLPAPVAPALNRTGDPG-----PGPSIOKTYDLYLTRYLHQLRS 56

Db 29 DAFNYQSCFMQDLFAHFPEVLFIHRTYNPGRKVLVTELVDPGVQVEGLARAV----- 82

QY 57 LAGTYLNYLGPFFNE-----PDFNPPRLGAETLPRATVDLEWRSINDKRLRLT 104

Db 83 -----YFAIPTNEDARGLAQMFQVKKFNP-----AWERVNTL----- 116

QY 105 QNYEAYSHLLCYLRLGNRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGVPLPQP 161

Db 117 ----VDPHFL--LLPTLTWEFPTAEVLLSAFICKFLQCKFYQL-----PLEQFVQR 162

QY 162 ----LPGTEPTWTPGAH-----SDFLQKMDDFWLLKELQTLW-----WRSKADF 202

Db 163 LLLSSLQSTMCSATAGNLKLVTLNLCIPSSRLPELHSHWLLND-RWLAWHRSRAQS 221

QY 203 NRLKKMKOPPAATAVLHGAHGF 225

Db 222 SRYFQSLEIMAHILSQFFGTTFP 244

RESULT 10

Q9MAUI PRELIMINARY; PRT; 332 AA.

ID Q9MAUI

AC Q9MAUI; 7.9%; Score 96.5; DB 11; Length 455;

DT 01-OCT-2000 (TREMELrel. 15, Created)

DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)

DE F13M7.8 protein.

GN F13M7.8.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

RN [1];

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Kwan A., Liu, S.,

Li J., Kremenetskaia I., Luros J., Araujo R., Au M., Bredel V.,

Buehler E., Conway A., Dewar K., Feng J., Kim C., Kurtz D., Li Y.,

Palm C., Shinn P., Sun H., Davis R., Ecker J., Federspiel N.,

Theologis A.,

RT "The sequence of BAC F13M7 from Arabidopsis thaliana chromosome 1.1";

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [2];

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Theologis A.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC004809; AAF40444.1; -.

DR InterPro; IPR002965; P-rich extensin.

DR PRINTS; PR01217; PRICHEXTENSIN.

```

SQ SEQUENCE 332 AA; 36793 MW; 18E8687141A070F4 CRC64;
Query Match
Best Local Similarity 26.4%; Score 92; DB 10; Length 332;
Matches 58; Conservative 24; Mismatches 72; Indels 66; Gaps 13;

QY 23 PAVPALNRTGDPGPGSPSTQKTYDLYLHQLRSLAGTYLNY-----64
D 113 PSVTAGNLGSGYP-PPSPF--TYDFGPEQRQWESLLQOFIREPNQIRPLRGLGLSPVG 169
QY 65 LGPPFNEPDPNPRLGAEATLPRATVDLEWVRSNDKRLTQNEYAYSHLLCYLRLGNRQA 124
D 170 LGPIRASPFQLOVRAPP--PTSILD--TSNRKAR-----SKGALAVVRG--RKV 215
QY 125 ATAELRSL-----AHFCTSLQGLSGIAGVMAALGYPLPQPLP---GTEPTWT 170
D 216 RITEGSSSLYSLGRSLKNGAHV-----GIQPRSGINK-----PLPKPLPVDLTITTSVP 266
QY 171 PGPAHSDFLQKMDDFWLLKELQTLWRSKADPNRLKKQKQ 210
D 267 DDPDESADKDEEAVKQL-----SEKDL--LKRHIE 298

RESULT 11
Q9JHE4 PRELIMINARY; PRT; 423 AA.
AC Q9JHE4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Cerebroside sulfotransferase.
GN GCST OR CST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Iyoy-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007645; BAB25160.1; -.
DR MGD; MGI:1858277; Gcst.
SQ SEQUENCE 423 AA; 48824 MW; 045303CFCE150A3 CRC64;

Query Match
Best Local Similarity 22.7%; Score 90; DB 11; Length 423;
Matches 55; Conservative 22; Mismatches 61; Indels 104; Gaps 12;

QY 44 YDLTRYLEHQLRSLAGTYLNY---LGPP-----FNEP-----72
D 204 YDPSSYNAHYLRNLLFFDLGYDSSLDPASPRVQEHILEVERRHFLVLLQYFDESILVLLQ 263
QY 73 -----DFNPPRLGAEATLPRAT-----VDLEWVRSNDKRLTQNEYA 109
D 264 ELLCWDLEDVLYFKLNARSDSPVRLSGELRYRATAWNLLDVLRYRHN--ASFWRKVEA 321
QY 110 YSHLLCYLRLGNRQA-ATAELRRS---LAHFCTSLQGLSGIA-----GVMAA 153
D 322 F-----GRERMAREVAELRQANEHMRHICIDGQAVGAERAIQDSAMQPQPLGIKSI 373
QY 154 LGYPLPQPLPGTEPTWTGPAHSDFLQKMDDFWLLKELQ-----TWLWRSKAD 202
D 374 LGYNLKKSI-----GPQHEQLCRGM-----LTPETQYLSDLGANLWVTKLWFLRDF 420
QY 203 NR 204
D 421 LR 422

RESULT 13
Q9S3Q9 PRELIMINARY; PRT; 771 AA.
ID Q9S3Q9
AC Q9S3Q9
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```

```

01-MAY-2000 (TREMBlrel. 13, Last sequence update)
01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Immunoreactive 87kd antigen PG92.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteriia; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;
OC Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "Porphyromonas gingivalis polypeptides and nucleic acids.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF175724; AAD51077.1; -.
DR TIGRFAMs; TIGR01180; aman2 put. 1.
SQ SEQUENCE 771 AA; 86667 MW; 75016BP66848C9B9 CRC64;

Query Match 7.3%; Score 89.5; DB 2; Length 771;
Best Local Similarity 25.6%; Pred. No. 9.4;
Matches 50; Conservative 16; Mismatches 48; Indels 81; Gaps 13;

QY 9 WGNLACLCITVL-----WHLPAVALNRTGDPGPGPSIQKTYDLYLRYLHQLRSLA----- 58
DB 549 YGM---LCPLLPGSFLTPDPKQGENFENPFHGSAYNYAFFVPHDIQGLARLMGGA 605

QY 59 -----GTYLNYLGGPPNPFDPNPRGAEATLPRATVDLEWRSNDKRLTQ 105
DB 606 KYFSERLQKVFDEGYV-----DPTNEPDIAYPVL-PSYFPK-----EAWR----- 644

QY 106 NYEAYSHLLCYLRLGNRAQTAEIRLSL--AHFCTSLQGLLGS-IAGVMAA-----LG- 155
DB 645 -----TQKLRELIDKHFCNAPNGLPGNDAGTMSAWLYVSMGLF 684

QY 156 YPLPQPLQGTPTWT 170
DB 685 YP---DCPGS-PIYT 695

RESULT 14
QY1664 PRELIMINARY; PRT; 733 AA.
AC QY1664;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE Hypothetical protein PA0454.
GN PA0454.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; Pubmed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004483; AAG03843.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 733 AA; 81538 MW; 580F9DBECB3909DB CRC64;

Query Match 7.3%; Score 89; DB 16; Length 733;
Best Local Similarity 26.6%; Pred. No. 9.8;
Matches 58; Conservative 26; Mismatches 66; Indels 68; Gaps 13;

QY 28 LNRTGDPGPGPSIQKTYDLYLRYLHQLRSLAGTYLNYLGGPPNPFDPNPRGAEATLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRLGNRAQTAEIRLSL 187
DB 230 LNRLGHRGPGKV-----SRYLKLYFIAQDVHERASSSHYPYNRLAFAEPHSDVLFRCQ 283
QY 78 RL-----GAETIPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRLGNRAQTAEIRLSL 131
DB 284 RLNLQOGKACQALARA-----TLRQPFDP-YA-----DRELALEDLQA 320
QY 132 SLAHF-----CTSLQGLLGSAGVMAALGYPLQPLQGTPTWTGPPAHSDFLQKMDDFWL 187
DB 321 SLEHLRQOSNPANKGLRLS-LGALAANLTLDRKLGA-----SNPDAIADQDSAL 371
QY 188 LKELQTLWMSAKD-FNRLKKQMPPAAAATLHLGAHG 224
DB 372 LDRSP-----RSLKDAFERLRQLTF-----TSLLFRHG 400

RESULT 15
QY1664 PRELIMINARY; PRT; 955 AA.
AC QY1664;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE WIZS.
GN WIZ OR WIZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Matsumoto K., Ishii N., Yoshida S., Shiosaka S., Wanaka A.,
RA Toyama M.;
RT "Molecular Cloning and Distinct Developmental Expression Pattern of
RT Spliced Forms of A Novel Zinc Finger Gene wiz in The Cerebellum.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AB012266; BAA32791.1; -.
DR MGB; MGI:1332638; WIZ.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 5.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 5.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 955 AA; 102777 MW; D4DD03A02BECCAAC CRC64;

Query Match 7.1%; Score 87; DB 11; Length 955;
Best Local Similarity 27.8%; Pred. No. 21;
Matches 35; Conservative 19; Mismatches 48; Indels 24; Gaps 7;

QY 22 LPAVALNRTGDPGPGPSIQKTYDLYLRYLHQLRSLAGTYLNYLGGPPNPFDPNPRGAEATLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRLGNRAQTAEIRLSL 73
DB 615 LPLSLASPRGPGAGPT-----QVPR--ELSLSPITGSPSAASYLGCVATKPELQEDR 667
QY 74 FNPPLRGAETLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRLGNRAQTAEIRLSL 130
DB 668 FLPAEVKATYIQTLELPFKA-KTLHEK-----TSHSSTACECLGCLYFENRKALASHAR 721
QY 131 RSLAHF 136
DB 722 AHLRQF 727

Search completed: January 27, 2003, 15:44:07
Job time : 28.5 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 1963.55 Seconds  
(without alignments)  
11812.771 Million cell updates/sec

Title: US-09-931-704-1  
Perfect score: 797  
Sequence: 1 attaagcttcgcgcgagcc.....tctctcttgcgtccccc 797

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rtd.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	797	100.0	797	6	AR002595	Sequence
2	797	100.0	797	6	AX392086	Sequence
3	797	100.0	797	9	AF176911	Homo sapi
4	793.4	99.5	881	6	AX205024	Sequence
5	793.4	99.5	881	6	AX205042	Sequence
6	776.4	97.4	1736	9	BC012939	Homo sapi
7	749.4	94.0	1689	9	AF172854	Homo sapi
8	680	85.3	680	9	AY049779	Homo sapi
9	669.4	84.0	819	6	AR002597	Sequence
10	669.4	84.0	819	6	AX392089	Sequence
11	669.4	84.0	819	10	AF176913	Mus muscu
12	594.4	74.6	1692	6	AX205060	Sequence
13	523.4	65.7	5087	6	AR002596	Sequence
14	523.4	65.7	5087	6	AX392088	Sequence
15	523.4	65.7	5087	9	AF176912	Homo sapi
16	523.4	65.7	135116	2	AP002437	Homo sapi
17	523.4	65.7	168567	9	AP003419	Homo sapi
18	521.8	65.5	169144	2	AC005849	Homo sapi
19	429.6	53.9	269355	2	AC109138	Homo sapi
20	269.4	33.8	283	6	AX202145	Sequence
21	86.4	10.8	63347	2	AC110526	Mus muscu
22	59.4	7.5	125020	9	AF429315	Homo sapi
23	57.8	7.3	125020	9	AF429315	Homo sapi
24	47.6	6.0	648	5	AF181480	Plethodon
25	47.6	6.0	648	5	AF181481	Plethodon
26	46.4	5.8	39552	2	AC097600	Rattus no
27	46.4	5.8	149120	2	AP005536	Oryza sat
28	46.4	5.8	187727	2	AC021142	Homo sapi
29	46.2	5.8	97826	9	AL591647	Human DNA
30	46.2	5.8	173609	9	AL451077	Human DNA
31	46	5.8	648	5	AF181482	Plethodon
32	45.6	5.7	1901	5	GGLAMB1	Chicken mRN
33	44.8	5.6	2265	9	BC009360	Homo sapi
34	44.8	5.6	16124	1	AX024384	Sequence
35	44.8	5.6	16124	6	AX024277	Sequence
36	44.8	5.6	68750	1	AF210843	Sorangium
37	44.8	5.6	68750	6	AR193029	Sequence
38	44.8	5.6	68750	6	AR199551	Sequence
39	44.8	5.6	68750	6	AR199559	Sequence
40	44.8	5.6	68750	6	AR199567	Sequence
41	44.8	5.6	68750	6	AR201097	Sequence
42	44.8	5.6	68750	6	AR208671	Sequence
43	44.4	5.6	648	5	AF181483	Plethodon
44	44	5.5	47958	2	AC091104	Homo sapi
45	44	5.5	168992	2	AC103568	Rattus no

ALIGNMENTS

RESULT 1  
LOCUS AR002595  
DEFINITION Sequence 1 from patent US 5741772.  
ACCESSION AR002595  
VERSION AR002595.1 GI:3964149  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 797)  
AUTHORS Chang, M.-s.  
TITLE Neurotrophic factor NNT-1  
JOURNAL Patent: US 5741772-A 1 21-APR-1998;  
FEATURES Location/Qualifiers

AR002595 797 bp DNA linear PAT 04-DEC-1998

```

source          1. .797
BASE COUNT      139 a 297 c 218 g 143 t
ORIGIN

Query Match      100.0%; Score 797; DB 6; Length 797;
Best Local Similarity 100.0%; Pred. No. 1.9e-152;
Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTAAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTTCGCGGAGAGGAG 60
DB 1 ATTTAAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTTCGCGGAGAGGAG 60
QY 61 CCGCACCCGCGCGCGAGCGCGCTCGCCCTCCCACTTCGCGGAGAGGAG 120
DB 61 CCGCACCCGCGCGCGAGCGCGCTCGCCCTCCCACTTCGCGGAGAGGAG 120
QY 121 TGTAGGCTGCTGTGACAGGTGCTTGGACCTTCCCTGAGTGCAGCTCTCAATCGCA 180
DB 121 TGTAGGCTGCTGTGACAGGTGCTTGGACCTTCCCTGAGTGCAGCTCTCAATCGCA 180
QY 181 CAGGGAGCCAGAGGCTTGGCGCTTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240
DB 181 CAGGGAGCCAGAGGCTTGGCGCTTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240
QY 241 AGCAACAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCGGCTTCAACG 300
DB 241 AGCAACAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCGGCTTCAACG 300
QY 301 AGCAGACTTCAACCTCCCGCTGGGCGAGAGCTCTGCCCGAGGCGCACTGTGTGACT 360
DB 301 AGCAGACTTCAACCTCCCGCTGGGCGAGAGCTCTGCCCGAGGCGCACTGTGTGACT 360
QY 361 TGGAGGTGTGGCGAAGGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCTTACA 420
DB 361 TGGAGGTGTGGCGAAGGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCTTACA 420
QY 421 GCCACCTTCTGTGTACTTGGTGGGCTTCAACCTGAGGCTGCGGCTGAGCTGCGGCT 480
DB 421 GCCACCTTCTGTGTACTTGGTGGGCTTCAACCTGAGGCTGCGGCTGAGCTGCGGCT 480
QY 481 GCAGCTGGGCGGCTTCTGACAGGCTTCCAGGCGCTTGGGCGAGCAATGGGCGGCTCA 540
DB 481 GCAGCTGGGCGGCTTCTGACAGGCTTCCAGGCGCTTGGGCGAGCAATGGGCGGCTCA 540
QY 541 TGGCAGCTCTGGGCTACCCACTGCGGCGAGGCTGCTGCGGAGCTGAAACCCACTTGGACTC 600
DB 541 TGGCAGCTCTGGGCTACCCACTGCGGCGAGGCTGCTGCGGAGCTGAAACCCACTTGGACTC 600

mat_peptide
171. .764
/product="Met at -27" 143 t
BASE COUNT      139 a 297 c 218 g 143 t
ORIGIN

human.
SOURCE           Homo sapiens
ORGANISM         Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE        1
AUTHORS          Senaldi, G.
TITLE            Methods and compositions for treating ige-related disease using nn
                  t-1 inhibitors
JOURNAL          Patent: WO 0215977-A 1 28-FEB-2002;
                  Amgen Inc. (US)
FEATURES         location/Qualifiers
source           1. .797
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  90. .767
                  /note="unnamed protein product"
                  /codon_start=1
                  /protein_id="CAD28985.1"
                  /db_xref="GI:19700575"
                  /translation="MDLRAGDSWGLACLTVLVHLPAVPALNRTGDPGPGPSIQITY
                  DLTRYLEHQLRSLAGTYLNLNQATAEALRSLAHFCTSLQGLGSIAGVMAALGYPUPQ
                  LTQNYEAYSHLLCYLRGLNRQATAEALRSLAHFCTSLQGLGSIAGVMAALGYPUPQ
                  PLPTETPTWTPGPAHSDFLQKMDDFMLLKLQTLWLSAKDFNRLKKMQPPAAAVTL
                  HLGAGHP"
                  171. .764
                  /product="Met at -27" 143 t

```

```

RESULT 2
LOCUS           AX392086
DEFINITION      Sequence 1 from Patent WO0215977.
ACCESSION       AX392086
VERSION         AX392086.1 GI:19700574
KEYWORDS        .

```



QY 601 CTGGCCCTGCCACAGTGAAGTCTCTCCAGAGATGGAGCACTTCTGGCTGTGTAAGGAGC 660  
 Db 601 CTGGCCCTGCCACAGTGAAGTCTCTCCAGAGATGGAGCACTTCTGGCTGTGTAAGGAGC 660  
 QY 661 TGCAGACCTGGCTGTGGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCGAGC 720  
 Db 661 TGCAGACCTGGCTGTGGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCGAGC 720  
 QY 721 CTCAGACGCTGCACTACCTGCACTTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780  
 Db 721 CTCAGACGCTGCACTACCTGCACTTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780  
 QY 781 CCTCTTGGCTCCCCCCC 797  
 Db 781 CCTCTTGGCTCCCCCCC 797

RESULT 3  
 AF176911  
 LOCUS Homo sapiens neurotrophin-1/B-cell stimulating factor-3 mRNA, PRI 04-OCT-1999  
 DEFINITION complete cds.  
 ACCESSION AF176911  
 VERSION AF176911.1 GI:6007640  
 KEYWORDS IL-6 family  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Senaldi,G., Varnum,B.C., Sarmiento,U., Lile,J., Starnes,C., Lile,J., Scully,S., Guo,J., Elliott,G., McNinch,J., Shaklee,C.L., Freeman,D., Manu,F., Simonet,W.S., Boone,T. and Chang,M.-S.  
 TITLE Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the IL-6 family  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)  
 MEDLINE 99432254  
 PUBMED 10500198

REFERENCE  
 AUTHORS Senaldi,G., Varnum,B., Sarmiento,U., Lile,J., Starnes,C., Scully,S., Guo,J., Elliott,G., McNinch,J., Freeman,D., Shaklee,C., Manu,F., Simonet,S., Boone,T. and Chang,M.-S.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-AUG-1999) Amgen, Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA

FEATURES  
 source  
 1..797  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 90..767  
 /codon\_start=1  
 /product="neurotrophin-1/B-cell stimulating factor-3"  
 /protein\_id="AA00991.1"  
 /db\_xref="GI:6007641"  
 /translation="MDLRAGDSWGLACITVHLPAVPAALNRDTPQPGPSIKTYDLTRYELHQLSLACTYLYNLPPEPNEPDPNPRIGAEFLPRATVDLEWRSLDKLRLTQYEAHSHLLCYLRGNLRQATAEALRSFAHCTSLQGLLGSYAGWALGYPLPQPLFGHEPTWPGFAHSDFLQKMDDFLLKLQTLWRSKADFNRLKKMQPPAAAVTLHIGAHGF"

BASE COUNT 139 a 297 c 218 g 143 t

ORIGIN  
 Query Match 100.0%; Score 797; DB 9; Length 797;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-152;  
 Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTAAGCTTCGCGGAGCGGGCTCGCCCTCCCACTCGCCAGCTCCGGGAGAGGAG 60  
 Db 1 ATTAAGCTTCGCGGAGCGGGCTCGCCCTCCCACTCGCCAGCTCCGGGAGAGGAG 60  
 QY 61 CCGCACC CGGCGGCCAGCCCGATGACCTCCGAGAGGGGACTCGTGGGGGA 120  
 Db 61 CCGCACC CGGCGGCCAGCCCGATGACCTCCGAGAGGGGACTCGTGGGGGA 120

QY 121 TGTAGCGTGTGACGGTGTCTGGCAACCTCCCTGCGAGTGCAGCTCTCAATCGCA 180  
 Db 121 TGTAGCGTGTGACGGTGTCTGGCAACCTCCCTGCGAGTGCAGCTCTCAATCGCA 180  
 QY 181 CAGGGACCCAGGGGCTGGCCCTCCATCCAGAAAAAATATGACCTCAACCGCTACCTGG 240  
 Db 181 CAGGGACCCAGGGGCTGGCCCTCCATCCAGAAAAAATATGACCTCAACCGCTACCTGG 240  
 QY 241 AGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTGGGCCCCCTTTTCAACG 300  
 Db 241 AGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTGGGCCCCCTTTTCAACG 300  
 QY 301 AGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCACTGTGTGACT 360  
 Db 301 AGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCACTGTGTGACT 360  
 QY 361 TGGAGGTGTGGCGAAGCTTCAATGACAAACTGCGGCTGACCCAGAACTTACGAGGCTTACA 420  
 Db 361 TGGAGGTGTGGCGAAGCTTCAATGACAAACTGCGGCTGACCCAGAACTTACGAGGCTTACA 420  
 QY 421 GCCACCTTCTGTGTACTTGGTGGCTCAACCGTCAAGGCTGACCTGTGAGTGGGCC 480  
 Db 421 GCCACCTTCTGTGTACTTGGTGGCTCAACCGTCAAGGCTGACCTGTGAGTGGGCC 480  
 QY 481 GCAGCTTGGGCCACTTCTGCACCAAGCTTCCAGGGGCTGGGGGAGCATTTGGGGGCTCA 540  
 Db 481 GCAGCTTGGGCCACTTCTGCACCAAGCTTCCAGGGGCTGGGGGAGCATTTGGGGGCTCA 540  
 QY 541 TGGCAGCTTGGGCTACCCACTGCCCCAGCGCTGCTGGGACTTGAACCCACTTGGGACTC 600  
 Db 541 TGGCAGCTTGGGCTACCCACTGCCCCAGCGCTGCTGGGACTTGAACCCACTTGGGACTC 600  
 QY 601 CTGGCCCTGCCACAGTGAAGTCTCTCAGAAAGATGACGACTTCTGCTGCTGTGAAAGGAGC 660  
 Db 601 CTGGCCCTGCCACAGTGAAGTCTCTCAGAAAGATGACGACTTCTGCTGCTGTGAAAGGAGC 660  
 QY 661 TGCAGACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCGAGC 720  
 Db 661 TGCAGACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCGAGC 720  
 QY 721 CTCAGACGCTGCACTACCTGCACTTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780  
 Db 721 CTCAGACGCTGCACTACCTGCACTTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780  
 QY 781 CCTCTTGGCTCCCCCCC 797  
 Db 781 CCTCTTGGCTCCCCCCC 797

RESULT 4  
 AX205024  
 LOCUS Sequence 1 from Patent WO0155172.  
 DEFINITION AX205024  
 ACCESSION AX205024  
 VERSION AX205024.1 GI:15394259  
 KEYWORDS human.  
 SOURCE Homo sapiens

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 881)  
 Elson,G., Gauchat,J.F., Plun-Favreau,H., Chevalier,S. and Gascan,H.  
 Isolated complex comprising a nnt-1 protein and in addition at least a c1f-1 protein and/or a scntff\_g(a) protein  
 Patent: WO 0155172-A 1 02-AUG-2001;  
 PIERRE FABRE MEDICAMENT (FR); INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM) (FR)  
 Location/Qualifiers  
 1..881  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 174..851

source  
 CDS

RESULT 5  
AX205042  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
CDS  
BASE COUNT 158 a 318 c 246 g 159 t  
ORIGIN  
Query Match 99.5%; Score 793.4; DB 6; Length 881;  
Best Local Similarity 99.9%; Pred. No. 1e-151;  
Matches 794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATTAAGCTTCGCGGAGCGCGCTCGCTCCCTCCACTCCGCGAGCTCCGCGAGAGGAG 60  
DB 85 ATTAAGCTTCGCGGAGCGCGCTCGCTCCCTCCACTCCGCGAGCTCCGCGAGAGGAG 144  
QY 61 CCGCACCCCG 120  
DB 145 CCGCACCCCG 204  
QY 121 TGTTAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAG 180  
DB 205 TGTTAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAG 264  
QY 181 CAGGGGACCCAGGGGCTTGGCGCTTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240  
DB 265 CAGGGGACCCAGGGGCTTGGCGCTTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 324  
QY 241 AGCACCAACTCCGAGCTTGGCTGGGAGCTATCTGAACTACTCTGGGCGCCCTTTCAAG 300  
DB 325 AGCACCAACTCCGAGCTTGGCTGGGAGCTATCTGAACTACTCTGGGCGCCCTTTCAAG 384  
QY 301 AGCAGACTTCAACCCCTCCCGCTCGGGGCGAGACTCTGCCAGGCGCACTGTGTGACT 360  
DB 385 AGCAGACTTCAACCCCTCCCGCTCGGGGCGAGACTCTGCCAGGCGCACTGTGTGACT 444  
QY 361 TGGAGGTGTGGCGAGCTTCAATGACAACTGCGGCTGACCCAGAACTACAGGGCTTACA 420  
DB 445 TGGAGGTGTGGCGAGCTTCAATGACAACTGCGGCTGACCCAGAACTACAGGGCTTACA 504  
QY 421 GCCACCTTCTGTGTGTACTTGGTGGCTTCAACCGCTCAGCTGCGCACTGTGTGAGCTG 480  
DB 505 GCCACCTTCTGTGTGTACTTGGTGGCTTCAACCGCTCAGCTGCGCACTGTGTGAGCTG 564  
QY 481 GCAGCTTGGCGCACTTCTTGCAACCGCTTCCAGGGGCTTCCAGGGGCTTCCAGGGGCT 540  
DB 565 GCAGCTTGGCGCACTTCTTGCAACCGCTTCCAGGGGCTTCCAGGGGCTTCCAGGGGCT 624

/note="NNT-1"  
/codon\_start=1  
/protein\_id="CAC60175.1"  
/db\_xref="GI:15394260"  
/translation="MDLRAGDSWGLACLTVLWHLPAVPAALNRTGDPGPGPSIOKTY  
DLTRYLEQLRLSLAGTYLNLGPPNEPDPNPRGLAETLPRATVDELVRSNDKLR  
LTQNYEAYSHLLCYLRGLNRQATLRLSLAHFCTSLQGLLGSIAGVAAALGYPQP  
PLPGTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLSAKDFNRLKKMQPPAAAVTL  
HLGAHGF"  
158 a 318 c 246 g 159 t  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 881)  
Eison, G. and Gauchat, J. F.  
Scsctfr/nnt-1 fusion protein  
Patent: WO 0155219-A 1 02-AUG-2001;  
PIERRE FABRE MEDICAMENT (FR)  
Location/Qualifiers  
1..881  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
174..851  
/note="NNT-1"  
/codon\_start=1  
/protein\_id="CAC60178.1"  
/db\_xref="GI:15394278"  
translation="MDLRAGDSWGLACLTVLWHLPAVPAALNRTGDPGPGPSIOKTY  
DLTRYLEQLRLSLAGTYLNLGPPNEPDPNPRGLAETLPRATVDELVRSNDKLR  
LTQNYEAYSHLLCYLRGLNRQATLRLSLAHFCTSLQGLLGSIAGVAAALGYPQP  
PLPGTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLSAKDFNRLKKMQPPAAAVTL  
HLGAHGF"  
158 a 318 c 246 g 159 t  
Query Match 99.5%; Score 793.4; DB 6; Length 881;  
Best Local Similarity 99.9%; Pred. No. 1e-151;  
Matches 794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATTAAGCTTCGCGGAGCGCGCTCGCTCCCTCCACTCCGCGAGCTCCGCGAGAGGAG 60  
DB 85 ATTAAGCTTCGCGGAGCGCGCTCGCTCCCTCCACTCCGCGAGCTCCGCGAGAGGAG 144  
QY 61 CCGCACCCCG 120  
DB 145 CCGCACCCCG 204  
QY 121 TGTTAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAG 180  
DB 205 TGTTAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAG 264  
QY 181 CAGGGGACCCAGGGGCTTGGCGCTTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240  
DB 265 CAGGGGACCCAGGGGCTTGGCGCTTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 324  
QY 241 AGCACCAACTCCGAGCTTGGCTGGGAGCTATCTGAACTACTCTGGGCGCCCTTTCAAG 300  
DB 325 AGCACCAACTCCGAGCTTGGCTGGGAGCTATCTGAACTACTCTGGGCGCCCTTTCAAG 384  
QY 301 AGCAGACTTCAACCCCTCCCGCTCGGGGCGAGACTCTGCCAGGCGCACTGTGTGACT 360  
DB 385 AGCAGACTTCAACCCCTCCCGCTCGGGGCGAGACTCTGCCAGGCGCACTGTGTGACT 444  
QY 361 TGGAGGTGTGGCGAGCTTCAATGACAACTGCGGCTGACCCAGAACTACAGGGCTTACA 420  
DB 445 TGGAGGTGTGGCGAGCTTCAATGACAACTGCGGCTGACCCAGAACTACAGGGCTTACA 504  
QY 421 GCCACCTTCTGTGTGTACTTGGTGGCTTCAACCGCTCAGCTGCGCACTGTGTGAGCTG 480  
DB 505 GCCACCTTCTGTGTGTACTTGGTGGCTTCAACCGCTCAGCTGCGCACTGTGTGAGCTG 564  
QY 481 GCAGCTTGGCGCACTTCTTGCAACCGCTTCCAGGGGCTTCCAGGGGCTTCCAGGGGCT 540  
DB 565 GCAGCTTGGCGCACTTCTTGCAACCGCTTCCAGGGGCTTCCAGGGGCTTCCAGGGGCT 624

541' TGGCAGCTCTGGCTACCCAGCTGCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 600  
 |||||  
 Db TGGCAGCTCTGGCTACCCAGCTGCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 684  
 |||||  
 Qy CTGGCCCTGCCACAGTCACTTCTCCAGAGATGAGAGACTTCTGGCTGCTGAAGGAGC 660  
 |||||  
 Db CTGGCCCTGCCACAGTCACTTCTCCAGAGATGAGAGACTTCTGGCTGCTGAAGGAGC 744  
 |||||  
 Qy TGCAGACTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 720  
 |||||  
 Db TGCAGACTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 804  
 |||||  
 Qy CTCAGAGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 780  
 |||||  
 Db CTCAGAGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 864  
 |||||  
 Qy CCTCTTGGCTCCCCC 795  
 |||||  
 Db CCTCTTGGCTCCCCC 879  
 |||||

RESULT 6  
 BC012939  
 LOCUS  
 DEFINITION  
 Homo sapiens, Similar to cardiotrophin-like cytokine;  
 neurotrophin-1/B-cell stimulating factor-3, clone MGC:21195  
 IMAGE:4453813, mRNA, complete cds.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 BC012939  
 Homo sapiens  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1736)  
 Direct Submission  
 Submitted (20-AUG-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [villalon@bcm.tmc.edu](mailto:villalon@bcm.tmc.edu)  
 Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,  
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
 Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LILNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 28 Row: B Column: 23  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 6007640.  
 Location/Qualifiers  
 1..1736  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="MGC:21195 IMAGE:4453813"  
 /tissue\_type="Kidney, hypernephroma"  
 /clone\_lib="NIH MGC\_89"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 73..750  
 /codon\_start=1  
 /product="Similar to cardiotrophin-like cytokine;

FEATURES  
 source

CDS

neurotrophin-1/B-cell stimulating factor-3"  
 /protein\_id="AAH12939.1"  
 /db\_xref="GI:15277895"

/translation="MDLRAGDSMGLACLCTVHLPAVPAALNRTGDPGPGSIQKTY  
 DLTRYLEHQSLAGTYINLYGPPFNPRLGAETLPRATVDLEWRLNDKLR  
 LTQYEAYSLLCYLRGLNQAAFAELRSLAHFCTSLQGLGSIAGVMAALGYPLPQ  
 PLPGTEPTWTPGPHSDFLQKMDDFLLKELQTLWRSADFNRLKKMQPAAAVTL  
 HLGAHGF"

BASE COUNT 371 a 546 c 454 g 365 t

Query Match 97.4%; Score 776.4; DB 9; Length 1736;  
 Best Local Similarity 99.9%; Pred. No. 2.7e-148;  
 Matches 777; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 18 GCCCGGCTGCGCTCCACATCCCGCAGGCTCCGGAGAGAGCGCACC CGCGGCC 77  
 |||||  
 Db 1 GCCCGGCTGCGCTCCACATCCCGCAGGCTCCGGAGAGAGCGCACC CGCGGCC 60  
 |||||  
 Qy 78 AGCCCCAGCCCCATGGACCTCCGAGCAGGGGACTCTGTTGGGGATGTTAGCGCTGTGC 137  
 |||||  
 Db 61 GGCCCCAGCCCCATGGACCTCCGAGCAGGGGACTCTGTTGGGGATGTTAGCGCTGTGC 120  
 |||||  
 Qy 138 ACGGTGCTTGGCACCCTCCCTGCGAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGGCT 197  
 |||||  
 Db 121 ACGGTGCTTGGCACCCTCCCTGCGAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGGCT 180  
 |||||  
 Qy 198 GGCCCTCCATCCAGAAACCTATGACCTCACCGCTACTCTGGAGCACCACCTCCGAGC 257  
 |||||  
 Db 181 GGCCCTCCATCCAGAAACCTATGACCTCACCGCTACTCTGGAGCACCACCTCCGAGC 240  
 |||||  
 Qy 258 TTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCT 317  
 |||||  
 Db 241 TTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCT 300  
 |||||  
 Qy 318 CCCCCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGACTTTGGAGGTGTGGCGAAGC 377  
 |||||  
 Db 301 CCCCCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGACTTTGAGAGGTGTGGCGAAGC 360  
 |||||  
 Qy 378 CTCAAATGACAAACTCGGGCTGACCCAGAACTACGAGGCTTACAGCACCCTTCTGTGTAC 437  
 |||||  
 Db 361 CTCAAATGACAAACTCGGGCTGACCCAGAACTACGAGGCTTACAGCACCCTTCTGTGTAC 420  
 |||||  
 Qy 438 TTGGCTGGGCTCAACCGTCCAGGCTGCGCTGAGCTGCGCGCCGAGCCCTGGCCACTTC 497  
 |||||  
 Db 421 TTGGCTGGGCTCAACCGTCCAGGCTGCGCTGAGCTGCGCGCCGAGCCCTGGCCACTTC 480  
 |||||  
 Qy 498 TGCACAGCTCCAGGGCTGCTGGGAGCATTTGCGGGGTCTATGCGAGCTCTGGGCTAC 557  
 |||||  
 Db 481 TGCACAGCTCCAGGGCTGCTGGGAGCATTTGCGGGGTCTATGCGAGCTCTGGGCTAC 540  
 |||||  
 Qy 558 CCACTGCCCCAGCGCTGCTGGGACTGAACCACTTGGACTCTGGCCCTGGCCCTGGCCACAGT 617  
 |||||  
 Db 541 CCACTGCCCCAGCGCTGCTGGGACTGAACCACTTGGACTCTGGCCCTGGCCCTGGCCACAGT 600  
 |||||  
 Qy 618 GACTTCTCCAGAGATGAGCAGCTTCTGGCTGCTGAAGGAGCTGCAGACTGGCTGTGG 677  
 |||||  
 Db 601 GACTTCTCCAGAGATGAGCAGCTTCTGGCTGCTGAAGGAGCTGCAGACTGGCTGTGG 660  
 |||||  
 Qy 678 CGCTGGCCCAAGGACTTCAACCGGCTCAAGAGAGATGAGCCTCCAGAGCTGCAGTC 737  
 |||||  
 Db 661 CGCTGGCCCAAGGACTTCAACCGGCTCAAGAGAGATGAGCCTCCAGAGCTGCAGTC 720  
 |||||  
 Qy 738 ACCCTGACCTGGGGCTCATGGCTTCTGACTTCTGACCTTCTGCTTCTGCTTCTGCTTCTGCT 795  
 |||||  
 Db 721 ACCCTGACCTGGGGCTCATGGCTTCTGACTTCTGACCTTCTGCTTCTGCTTCTGCTTCTGCT 778  
 |||||

RESULT 7  
 AF172854  
 LOCUS  
 DEFINITION  
 Homo sapiens  
 accession AF172854

AF172854  
 Homo sapiens  
 cardiotrophin-like cytokine  
 mRNA  
 linear  
 PRI 10-SEP-1999  
 complete cds.



Db 121 TCAGAAACCTATGACCTTACCGCTACTGGAGCAACCACTCCGACCTTGGCTGGGA 180  
 QY 268 CTAATCTGAATACCTTGGGCCCCCTTTCAAGGAGCAGACTTCAACCTCCCGCTGG 327  
 Db 181 CTAATCTGAATACCTTGGGCCCCCTTTCAAGGAGCAGACTTCAACCTCCCGCTGG 240  
 QY 328 GGGCAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGGGAGGCTCAATGACA 387  
 Db 241 GGGCAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGGGAGGCTCAATGACA 300  
 QY 388 AACTCGGGCTGACCCAGAGCTACGAGGCTACAGGCTCAGAGCCACTTCTGTTACTTGGCTGGCC 447  
 Db 301 AACTCGGGCTGACCCAGAGCTACGAGGCTCAGAGCCACTTCTGTTACTTGGCTGGCC 360  
 QY 448 TCAACCGTCAAGCTGCTGAGCTGCGCGGCTCAGAGCTTGGCCACTTCTGCAACAGCC 507  
 Db 361 TCAACCGTCAAGCTGCTGAGCTGCGCGGCTCAGAGCTTGGCCACTTCTGCAACAGCC 420  
 QY 508 TCCAGGGCTCTGCGGAGCATTGGGGGCTCATGGAGCTTGGGCTACCACTGCCCC 567  
 Db 421 TCCAGGGCTCTGCGGAGCATTGGGGGCTCATGGAGCTTGGGCTACCACTGCCCC 480  
 QY 568 AGCCGCTGCTGGGACTGAACCCACTTGGACTGCGCGGCTCAGAGCTTGGCCACTTCTGCAACAGCC 627  
 Db 481 AGCCGCTGCTGGGACTGAACCCACTTGGACTGCGCGGCTCAGAGCTTGGCCACTTCTGCAACAGCC 540  
 QY 628 AGAAGTGGAGACTTCTGGCTGCTGAAGGAGCTGCAAGCTGCTGGGCTCGGCCA 687  
 Db 541 AGAAGTGGAGACTTCTGGCTGCTGAAGGAGCTGCAAGCTGCTGGGCTCGGCCA 600  
 QY 688 AGGACTTCAACCGCTCAAGAGAGATGCAAGCTTCCAGAGCTGCAAGCTTCCAGCTGCACTGCACT 747  
 Db 601 AGGACTTCAACCGCTCAAGAGAGATGCAAGCTTCCAGAGCTGCAAGCTTCCAGCTGCACTGCACT 660  
 QY 748 TGGGGCTCATGGCTTCTGA 767  
 Db 661 TGGGGCTCATGGCTTCTGA 680

RESULT 9  
 LOCUS AR002597 819 bp DNA linear PAT 04-DEC-1998  
 DEFINITION Sequence 4 from patent US 5741772.  
 ACCESSION AR002597  
 VERSION AR002597.1 GI:3964151  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 819)  
 AUTHORS Chang, M.-s.  
 TITLE Neurotrophic factor NNT-1  
 JOURNAL Patent: US 5741772-A 4 21-APR-1998;  
 FEATURES Location/Qualifiers

source 1. .819 /organism="unknown"

BASE COUNT 156 a 288 c 218 g 157 t

ORIGIN

Query Match 84.0%; Score 669.4; DB 6; Length 819;  
 Best Local Similarity 92.0%; Pred. No. 1.9e-126;  
 Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

QY 1 ATTAAAGCTTCGCGAGCGGGCTCGGCTCCCACTCGCCAGCTCCGGAGAGGAG 60  
 Db 5 ATTAAAGCTTCGCGAGCGGGCTCGGCTCCCACTCGCCAGCTTCGGAGAGGAG 64

QY 61 CCGCAACCGCGCGGCCAG-CCCCAGCCCATGGACTCCGAGCAGGGGACTCGTGGGG 119  
 Db 65 CCGCGCGCGCGGCCAGCCCCAGCCCATGGACTCCGAGCAGGGGACTCGTGGGGG 124

QY 120 ATGTTAGCTGCTGTGACGGTGTCTGGCACTTCCCTGCTGAGTGCAGCTCTCAATCGC 179

Db 125 ATGTTAGCTGCTGTGACGGTGTGTGGACCTCCCTGCGAGTGCAGCTCTTAATCGC 184  
 QY 180 ACAGGGAGCCAGGCGCTGGCCCCCTCATCCAGAAAACTATGACCTACCCGCTACCTG 239  
 Db 185 ACAGGAGATCCAGGCGCTGGCCCCCTCATCCAGAAAACTATGACCTACCCGCTACCTG 244  
 QY 240 GAGCACCAGCTCCGAGCTTGGCTGGAGCTTCTGAACCTACTGCGGCCCCCTTTCAAC 299  
 Db 245 GAGCATCACTCCGAGCTTAGCTGGACCTACCTGAACTACTCTGGGGCCCCCTTTCAAC 304  
 QY 300 GAGCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGSCCACTGTTGAC 359  
 Db 305 GAGCCTGACTTCAATCTCTCGACTGGGGGAGAGAACTCTGCCAGGSCCACTGTTGAC 364  
 QY 360 TTGAGGTGTGGGAGGCTCAATGACAAATGCGGCTGACCCAGAACTACGAGGCTTAC 419  
 Db 365 TTGAGGTGTGGGAGGCTCAATGACAGAGCTGCGGCTGACCCAGAACTATGAGGCTTAC 424  
 QY 420 AGCCAGCTTCTGTGTTACTTGGCTGGCTCAACCTGAGGCTGCCACTGCTGAGTGGCC 479  
 Db 425 AGTCACTCTCTGTGTTACTTGGCTGGCTCAACCTGAGGCTGCCACAGCTGAATCCGA 484  
 QY 480 CGCAGCTTGGCCCCACTTCTGCAACAGGCTCCAGGCGCTTGGGGAGCAATTGCGGGCTG 539  
 Db 485 CGTAGCTTGGCCCCACTTCTGTACAGGCTCCAGGCGCTTGGGGAGCAATTGCGGGCTG 544  
 QY 540 ATGCGAGCTTGGGCTACCCACTGCCAGCGCTGCTGGGAGCTGAACCCACTTGGACT 599  
 Db 545 ATGCGAGCTTGGCTTACCCACTGCCAGCGCTTGGGGAGCAATTGCGGGCTG 604  
 QY 600 CTTGGCCCTGCCCCAGTACTTCTCCAGAGATGAGCACTTCAACCGGCTCAAGAGAGATGCG 719  
 Db 605 CTTGGCCCTGCCCCAGTACTTCTCCAGAGATGAGCACTTCAACCGGCTTAAAGAGAGATGCG 724  
 QY 720 CTTCCAGCAGCTGCGCTCAGCTGACCTGGGGCTCATGGCTTCTGACTTCTGACTT 778  
 Db 725 CTTCCAGCAGCTTCACTGCTGCTGCTGAGGAGCACTGCTGCTGACTTCTGACTT 783

RESULT 10  
 LOCUS AX392089 819 bp DNA linear PAT 23-MAR-2002  
 DEFINITION Sequence 4 from Patent WO0215977.  
 ACCESSION AX392089  
 VERSION AX392089.1 GI:19700577  
 KEYWORDS  
 SOURCE Mus sp.  
 ORGANISM Mus sp.

REFERENCE 1  
 AUTHORS Senaldi, G.  
 TITLE Methods and compositions for treating ige-related disease using nn

JOURNAL Patent: WO 0215977-A 4 28-FEB-2002;  
 Amgen Inc. (US)

FEATURES Location/Qualifiers  
 source 1. .819 /organism="Mus sp."  
 CDS 95..772 /db\_xref="taxon:10095"  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAD28986.1"  
 /db\_xref="GI:19700578"  
 /translation="MDLRAGDSWGLACITVHLPAVPALNRTGDPGPGPSIOKTY  
 DLTRYLHQRLSLAGTYLNYLGPFPNPFDPNPLGAEETLPRTATVNLFWKSLNDRLR  
 LTQYEAYSHLCLYLRLGNLRQAELRRSLAHFCTSLQGLLGSITAGVMTLGYPLPQ

PLPGETPAWAPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAASVTL  
HLEAHGF"  
sig\_peptide 95..175  
mat\_peptide 176..769  
/product="unnamed"  
BASE COUNT 156 a 288 c 218 g 157 t  
ORIGIN

Query Match 84.0%; Score 669.4; DB 6; Length 819;  
Best Local Similarity 92.0%; Pred. No. 1.9e-126;  
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
QY 1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 60  
Db 5 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 64  
QY 61 CCGACCCGCGCGCCAG-CCCGAGCCCATGACCTCCGAGCAGGAGGACTCGTGGGG 119  
Db 65 CCGCGCCGCGCGCGCCAGCCCGAGCCCATGACCTCCGAGCAGGAGGACTCGTGGGG 124  
QY 120 ATGTTAGCGCTGTGACCGTGTCTGCGACCTCCCTGCGAGTGCCAGCTCTCAATCGC 179  
Db 125 ATGTTAGCTGTGACCGTGTCTGCGACCTCCCTGCGAGTGCCAGCTCTCAATCGC 184  
QY 180 ACAGGGAGCCAGGCGCTGCGCTCGGACCTATCTGAATCTACTGGGCCCCCTTTCAAC 239  
Db 185 ACAGGAGATCCAGGCGCTGCGCTCGGACCTATCTGAATCTACTGGGCCCCCTTTCAAC 244  
QY 240 GAGCACCCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 299  
Db 245 GAGCACCCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 304  
QY 300 GAGCACCCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 359  
Db 305 GAGCACCCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 364  
QY 360 TTGGAGGTGTGGCGAAGCTCTAATGACAACTCCGCTGACCCAGAACTACGAGGCGCTAC 419  
Db 365 TTGGAAGTGTGGCGAAGCTCTAATGACAACTCCGCTGACCCAGAACTATGAGGCGTAC 424  
QY 420 AGCCACCTTCGCTGTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCCG 479  
Db 425 AGTCACTTCGCTGTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCCG 484  
QY 480 CCGACCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 539  
Db 485 CCGACCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 544  
QY 540 ATGGCAGCTTCGGCTTACCCACTGCGCCCGCGCTGCTGGGAGTGAACCTTGGACT 599  
Db 545 ATGGCAGCTTCGGCTTACCCACTGCGCCCGCGCTGCTGGGAGTGAACCTTGGACT 604  
QY 600 CTGGCCCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 659  
Db 605 CTGGCCCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 664  
QY 660 CTGACAGCTTCGCTGTGGGCGCTCGCCAGGACTTCAACCGCTCAGAGAGGAGTGCAG 719  
Db 665 CTGACAGCTTCGCTGTGGGCGCTCGCCAGGACTTCAACCGCTCAGAGAGGAGTGCAG 724  
QY 720 CTTCAGGAGCTGCAGTCACTTCGCTGCGCTCATGGCTTCTGACTTCTGACCTT 778  
Db 725 CTTCAGGAGCTGCAGTCACTTCGCTGCGCTCATGGCTTCTGACTTCTGACCTT 783

RESULT 11

AF176913

LOCUS

DEFINITION

AF176913

ACCESSION

VERSION

KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## CDS

Mus musculus.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 819)

Senaldi, G., Varnum, B.C., Sarmiento, U., Starnes, C., Lile, J.,

Scully, S., Guo, J., Elliott, G., McNinch, J., Freeman, D., Boone, T., and Chang, M.-S.

Freeman, D., Manu, F., Simonet, S., Boone, T., and Chang, M.-S.

Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the

IL-6 family

Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)

99432254

10500198

2 (bases 1 to 819)

Senaldi, G., Varnum, B., Sarmiento, U., Lile, J., Starnes, C.,

Scully, S., Guo, J., Elliott, G., McNinch, J., Freeman, D., Boone, T., and Chang, M.-S.

Manu, F., Simonet, S., Boone, T., and Chang, M.-S.

Direct Submission

Submitted (11-AUG-1999) Amgen, Inc., One Amgen Center Drive,

Thousand Oaks, CA 91320, USA

Location/Qualifiers

1..819

/organism="Mus musculus"

/db\_xref="taxon:10090"

95..772

/codon\_start=1

/product="neurotrophin-1/B-cell stimulating factor-3"

/protein\_id="AAF00993.1"

/db\_xref="GI:6007645"

/translation="MDLRAGDSWGLACLTVLWHLPAVPAALNRTPGPGPSIQKTY

DUTRYEHLRLSLAGTYLNYLPPNEPDPNPRLGAETLPATVNLVWRSNDRLR

LTQNYEAYSHLLCYLRGNRQATAEELRSLAHFCTSLQGLGSIAGVMATLGYPLPQ

PLPGETPAWAPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAASVTL

HLEAHGF"

BASE COUNT 156 a 288 c 218 g 157 t

## ORIGIN

Query Match 84.0%; Score 669.4; DB 10; Length 819;  
Best Local Similarity 92.0%; Pred. No. 1.9e-126;  
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

QY 1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 60  
Db 5 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 64  
QY 61 CCGACCCGCGCGCCAG-CCCGAGCCCATGACCTCCGAGCAGGAGGACTCGTGGGG 119  
Db 65 CCGCGCCGCGCGCGCCAGCCCGAGCCCATGACCTCCGAGCAGGAGGACTCGTGGGG 124  
QY 120 ATGTTAGCGCTGTGACCGTGTCTGCGACCTCCCTGCGAGTGCCAGCTCTCAATCGC 179  
Db 125 ATGTTAGCTGTGACCGTGTCTGCGACCTCCCTGCGAGTGCCAGCTCTCAATCGC 184  
QY 180 ACAGGGAGCCAGGCGCTGCGCTCGGACCTCCCACTCCAGAACTATGAGCTACCCGCTACCTG 239  
Db 185 ACAGGAGATCCAGGCGCTGCGCTCGGACCTCCCACTCCAGAACTATGAGCTACCCGCTACCTG 244  
QY 240 GAGCACCCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 299  
Db 245 GAGCACCCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 304  
QY 300 GAGCACCCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 359  
Db 305 GAGCACCCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCAGCTCCCGGAGAGGAG 364  
QY 360 TTGGAGGTGTGGCGAAGCTCTAATGACAACTCCGCTGACCCAGAACTACGAGGCGCTAC 419  
Db 365 TTGGAAGTGTGGCGAAGCTCTAATGACAACTCCGCTGACCCAGAACTATGAGGCGTAC 424  
QY 420 AGCCACCTTCGCTGTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCCG 479  
Db 425 AGTCACTTCGCTGTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCCG 484

QY 480 CGCAGCTGGCCCACTTCTGACAGCCTCCAGGGCTCTGCGGAGCACTTGGCGGCTC 539  
Db 485 CGTAGCTGGCCCACTTCTGACAGCCTCCAGGGCTCTGCGGAGCACTTGGCGGCTC 544  
QY 540 ATGGCAGCTCTGGCTACCACTGCCCCAGCGCTGCTGGGAGCACTTGGGACT 599  
Db 545 ATGGCAGCTTGGCTACCACTGCCCCAGCGCTCTGCGGAGCACTTGGGACT 604  
QY 600 CTGGCCCTGGCCCACTGAGTACTTCTCCAGAGATGGAGCACTTCTGGGCTGCTGAAGGAG 659  
Db 605 CTGGCCCTGGCCCACTGAGTACTTCTCCAGAGATGGAGCACTTCTGGGCTGCTGAAGGAG 664  
QY 660 CTGAGACTGCTGGCTGGCCCACTGAGTACTTCAACCGGCTCAAGAGAGATGAG 719  
Db 665 CTGAGACTGCTGGCTGGCCCACTGAGTACTTCAACCGGCTCAAGAGAGATGAG 724  
QY 720 CTCCAGAGCTGAGTCACTGACCTGACCTGGGGCTCATGGCTCTGAGCTTCTGAGCTT 778  
Db 725 CTCCAGAGCTTCACTGACCTGACCTGGAGGACATGGTTTCTGACCTCTGACCT 783

RESULT 12  
LOCUS AX205060 1692 bp DNA linear PAT 30-AUG-2001  
DEFINITION Sequence 19 from Patent WO0155219.  
ACCESSION AX205060  
VERSION AX205060.1 GI:15394299  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 1692)  
AUTHORS Elson, G. and Gauchat, J.F.  
TITLE Scentfr/nmt-1 fusion protein  
JOURNAL Patent: WO 0155219-A 19 02-AUG-2001;  
PIERRE FABRE MEDICAMENT (FR)  
FEATURES  
source 1. .1692  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
1. .1692  
/notes="proteine de fusion"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC60181.1"  
/db\_xref="GI:15394300"  
/translation="MAAPVFWACCAVLAIAAAVVAQHRSPQEAHPVQYERLGSVDVTL  
PGTANWDAATVRVNGTDLAPDLLGSLVHLGLHSLGLYACPHRDSWHLRHOVL  
LHVGPPREPVLSCRSNTYKGYFYSWHLPTPTYPNTFNTVTLHSGSKIMVCEKDAL  
KNRCHIRYMHLESTTKYKVISVSNALGNATAITFDEFTIVKPPPPENVARPPSPN  
PRRLEVTWTPSTWDPSPFPLKFLRPLLDQWQHVLSGDTAHTITDAYAKKY  
IIQVAKONEICTWSDMSVAHATPWTFRHLTEAQAETTTSTSLAPPPTTKI  
CDLEGGSGGGGGGSLNRTGDPGPSIQTYDLTRYLHQLRSLAGTYLNY  
LGPPNEPDPNPRLGAETLPATVDLEVRSLNDKRLTQNYEAYSHLLCYRLGNLR  
QAATALRRSLAHFETLSLQGLGSIAGVMAALGYPLPOLPTEPTETWTPGPHSDFLQ  
KMDDFLLKLELTWLRSAKDFNRLKKMQPPAAVTLHLGAGHFEQLISEEDL"

BASE COUNT 337 a 582 c 470 g 303 t  
ORIGIN  
Query Match 74.6%; Score 594.4; DB 6; Length 1692;  
Best Local Similarity 98.2%; Pred. No. 3.2e-111;  
Matches 601; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 171 CTCATCGCAGAGGACCCAGGGCTGGCCCTCCATCCAGAAAACCTATGACCTCACC 230  
Db 1066 CTCATCGCAGAGGACCCAGGGCTGGCCCTCCATCCAGAAAACCTATGACCTCACC 1125  
QY 231 CGCTACCTGGAGACA CAATCCGCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCC 290  
Db 1126 CGCTACCTGGAGACA CAATCCGCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCC 1185  
QY 291 CTTTTCACGAGCCAGACTTGAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCC 350

Db 1186 CTTTTCAAGCAGCCAGACTTCAACCCCTCCCGCTGGGGGAGAGACTCTGCCAGGGCC 1245  
QY 351 ACTGTGACTTGGAGGTGTGGGAAAGCCTCAATGACAAACTCGGGCTGACCCAGAACTAC 410  
Db 1246 ACTGTGACTTGGAGGTGTGGGAAAGCCTCAATGACAAACTCGGGCTGACCCAGAACTAC 1305  
QY 411 GAGGCTACAGCCACTTCTGTGTTACTTGTGGCTCAACCGTCAAGCTGCTGCTGCT 470  
Db 1306 GAGGCTACAGCCACTTCTGTGTTACTTGTGGCTCAACCGTCAAGCTGCTGCTGCT 1365  
QY 471 GAGCTGCGCCGAGCCTGCGCCCACTTCTGACACCGCTCCAGGGCTGCTGGGAGCACTT 530  
Db 1366 GAGCTGCGCCGAGCCTGCGCCCACTTCTGACACCGCTCCAGGGCTGCTGGGAGCACTT 1425  
QY 531 GCGGCGTATGGAGCTCTGGGCTTACCCACTGCCCCAGCGCTGCTGGGAGCACTTCTGGCTG 590  
Db 1426 GCGGCGTATGGAGCTCTGGGCTTACCCACTGCCCCAGCGCTGCTGGGAGCACTTCTGGCTG 1485  
QY 591 ACTTGGACTCTGCGCCCTGCGCCCACTGAGTACTTCTCCAGAAAGATGAGCACTTCTGGCTG 650  
Db 1486 ACTTGGACTCTGCGCCCTGCGCCCACTGAGTACTTCTCCAGAAAGATGAGCACTTCTGGCTG 1545  
QY 651 CTGAAGAGCTGACAGACTGCTGCTGGGCTGCGGCTCGGCAAGGACTTCAACCGGCTCAAGAG 710  
Db 1546 CTGAAGAGCTGACAGACTGCTGCTGGGCTGCGGCTCGGCAAGGACTTCAACCGGCTCAAGAG 1605  
QY 711 AAGATGAGCTTCCAGAGCTGAGTCACTGCACTTGGGGCTCATGGCTTCTGACTT 770  
Db 1606 AAGATGAGCTTCCAGAGCTGAGTCACTGCACTTGGGGCTCATGGCTTCTGAGCAG 1665  
QY 771 CTGACCTTCTCC 782  
Db 1666 AAGCTGATCTCC 1677

RESULT 13  
LOCUS AR002596 5087 bp DNA linear PAT 04-DEC-1998  
DEFINITION Sequence 3 from patent US 5741772.  
ACCESSION AR002596  
VERSION AR002596.1 GI:3964150  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5087)  
AUTHORS Chang, M.-s.  
TITLE Neurotrophic factor NNT-1  
JOURNAL Patent: US 5741772-A 3 21-APR-1998;  
FEATURES Location/Qualifiers  
source 1. .5087  
/organism="unknown"  
BASE COUNT 992 a 1746 c 1191 g 1158 t  
ORIGIN  
Query Match 65.7%; Score 523.4; DB 6; Length 5087;  
Best Local Similarity 99.8%; Pred. No. 8e-97;  
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 271 ATCTGAACCTACCTGGGCCCCCTTTCAACAGCGAGACTTCAACCCCTCCCGCTGGGG 330  
Db 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACAGCGAGACTTCAACCCCTCCCGCTGGGG 3422  
QY 331 CAGAGACTCTGCGCCAGGGCCACTGTTGACTTGGAGGTGTGGGAGCCTCAATGACAAAC 390  
Db 3423 CAGAGACTCTGCGCCAGGGCCACTGTTGACTTGGAGGTGTGGGAGCCTCAATGACAAAC 3482  
QY 391 TCGGCTGACCCAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGGCTGGGCTCA 450  
Db 3483 TCGGCTGACCCAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGGCTGGGCTCA 3542  
QY 451 ACCGTGAGCTGCCACTGCTGAGCTGCGCGAGCTGCGCCCACTTCTGACAGCCTCC 510



3543 ACCGTGAGGCTGCACTGCTGAGCTGCGCGCAGCCTGGCCCACTTCTGACACGAGCTCC 3602

511 AGGGCTGCTGGCAGCAATTCGGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAGC 570

3603 AGGGCTGCTGGCAGCAATTCGGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAGC 3662

571 CGCTGCTGGGACTGAACCCACTTGGACTCCTGCCCTGGCCCAAGTGAATTCCTCCAGA 630

3663 CGCTGCTGGGACTGAACCCACTTGGACTCCTGCCCTGGCCCAAGTGAATTCCTCCAGA 3722

631 AGATGACGACTTCTGGCTGCTGAGAGCTGCGAGCTGCGCTGCGGCTCGGCCAAGG 690

3723 AGATGACGACTTCTGGCTGCTGAGAGCTGCGAGCTGCGCTGCGGCTCGGCCAAGG 3782

691 ACTTCAACCGGCTCAAGAAGATGACAGCTTCCAGCAGCTGCAGTCACTGACCTGCTGG 750

3783 ACTTCAACCGGCTCAAGAAGATGACAGCTTCCAGCAGCTGCAGTCACTGACCTGCTGG 3842

751 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTCCCCC 795

3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTCCCCC 3887

RESULT 15

AF176912 5087 bp DNA linear PRI 04-OCT-1999

LOCUS AF176912

DEFINITION Homo sapiens neurotrophin-1/B-cell stimulating factor-3 gene, complete cds.

ACCESSION AF176912

VERSION AF176912.1 GI:6007642

KEYWORDS

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 5087)

AUTHORS Senaldi,G., Varnum,B.C., Sarmiento,U., Lile,J., Starnes,C., Lile,J., Scully,S., Guo,J., Elliott,G., McNinch,J., Shukle,C.L., Freeman,D., Manu,F., Simonet,W.S., Boone,T. and Chang,M.S.

TITLE Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the IL-6 family

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)

MEDLINE 99432254

PUBMED 10500198

REFERENCE 2 (bases 1 to 5087)

AUTHORS Senaldi,G., Varnum,B., Sarmiento,U., Lile,J., Starnes,C., Scully,S., Guo,J., Elliott,G., McNinch,J., Freeman,D., Manu,F., Simonet,S., Boone,T. and Chang,M.-S.

TITLE Direct Submission

JOURNAL Submitted (11-AUG-1999) Amgen, Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA

FEATURES

source 1..5087

organism="Homo sapiens"

db\_xref="taxon:9606"

join(120..135,1368..1534,3365..>3859)

/product="neurotrophin-1/B-cell stimulating factor-3"

join(120..135,1368..1534,3365..3859)

/codon\_start=1

/product="neurotrophin-1/B-cell stimulating factor-3"

/protein\_id="AAFP00992.1"

/db\_xref="GI:6007643"

/translation="MDLRAGDSWGLACLCTVLMHLPVAPNLRNTRGDPGPGSIQKTYDLRYLEHQLRSLAGTYLVNLYLGPFPNDFNPRGLAETLPRATVDLEWVRSLNDKLR LTONYEAYSHLLCYLRNROATLRSLAHFCTSLQGLLGSIAQYMAALGYPLPQ PLQTEPTWTPGPAHSDFLQKDDFWLLKELQTLWRSKAFNRLKKMQPPAAAVTL HLGAHGF"

BASE COUNT 992 a 1746 c 1191 g 1158 t

ORIGIN

Query Match 65.7%; Score 523.4; DB 9; Length 5087;

Best Local Similarity 99.8%; Pred. No. 8e-97;

Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCAGCACTTCAACCTCCCGGCTGGGG 330

Db 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCAGCACTTCAACCTCCCGGCTGGGG 3422

QY 331 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 390

Db 3423 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 3482

QY 391 TGGCGGTGACCCAGAACTACAGGCGCTACAGCCACTTCTGTGTTACTTGGTGGCGCTCA 450

Db 3483 TGGCGGTGACCCAGAACTACAGGCGCTACAGCCACTTCTGTGTTACTTGGTGGCGCTCA 3542

QY 451 ACCGTGAGGTGACCTGCTGAGTGGCGGCGAGCCTGGCCCACTTGTGACACGACCTCC 510

Db 3543 ACCGTGAGGTGACCTGCTGAGTGGCGGCGAGCCTGGCCCACTTGTGACACGACCTCC 3602

QY 511 AGGCGCTGCTGGGAGCAATTCGGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAGC 570

Db 3603 AGGCGCTGCTGGGAGCAATTCGGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAGC 3662

3543 ACCGTGAGGCTGCACTGCTGAGCTGCGCGCAGCCTGGCCCACTTCTGACACGAGCTCC 3602

511 AGGGCTGCTGGCAGCAATTCGGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAGC 570

3603 AGGGCTGCTGGCAGCAATTCGGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAGC 3662

571 CGCTGCTGGGACTGAACCCACTTGGACTCCTGCCCTGGCCCAAGTGAATTCCTCCAGA 630

3663 CGCTGCTGGGACTGAACCCACTTGGACTCCTGCCCTGGCCCAAGTGAATTCCTCCAGA 3722

631 AGATGACGACTTCTGGCTGCTGAGAGCTGCGAGCTGCGCTGCGGCTCGGCCAAGG 690

3723 AGATGACGACTTCTGGCTGCTGAGAGCTGCGAGCTGCGCTGCGGCTCGGCCAAGG 3782

691 ACTTCAACCGGCTCAAGAAGATGACAGCTTCCAGCAGCTGCAGTCACTGACCTGCTGG 750

3783 ACTTCAACCGGCTCAAGAAGATGACAGCTTCCAGCAGCTGCAGTCACTGACCTGCTGG 3842

751 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTCCCCC 795

3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTCCCCC 3887

RESULT 14

AX392088 5087 bp DNA linear PAT 23-MAR-2002

LOCUS AX392088

DEFINITION Sequence 3 from Patent WO0215977.

ACCESSION AX392088

VERSION AX392088.1 GI:19700576

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1

AUTHORS Senaldi,G.

TITLE Methods and compositions for treating ige-related disease using nn t-1 inhibitors

JOURNAL Patent: WO 0215977-A 3 28-FEB-2002; Amgen Inc. (US)

FEATURES

source 1..5087

organism="Homo sapiens"

db\_xref="taxon:9606"

misc\_feature 137..138

/note="INTERVENING UNSEQUENCED REGION OF >1KB"

BASE COUNT 992 a 1746 c 1191 g 1158 t

ORIGIN

Query Match 65.7%; Score 523.4; DB 6; Length 5087;

Best Local Similarity 99.8%; Pred. No. 8e-97;

Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCAGCACTTCAACCTCCCGGCTGGGG 330

Db 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCAGCACTTCAACCTCCCGGCTGGGG 3422

QY 331 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 390

Db 3423 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 3482

QY 391 TGGCGGTGACCCAGAACTACAGGCGCTACAGCCACTTCTGTGTTACTTGGTGGCGCTCA 450

Db 3483 TGGCGGTGACCCAGAACTACAGGCGCTACAGCCACTTCTGTGTTACTTGGTGGCGCTCA 3542

QY 451 ACCGTGAGGTGACCTGCTGAGTGGCGGCGAGCCTGGCCCACTTGTGACACGACCTCC 510

Db 3543 ACCGTGAGGTGACCTGCTGAGTGGCGGCGAGCCTGGCCCACTTGTGACACGACCTCC 3602

QY 511 AGGCGCTGCTGGGAGCAATTCGGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAGC 570

Db 3603 AGGCGCTGCTGGGAGCAATTCGGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAGC 3662



Search completed: February 1, 2003, 17:34:10  
Job time : 1980.55 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 1618.85 Seconds  
(without alignments)  
7973.437 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 797

Sequence: 1 attaaagcttcgcggagcc.....tctctcttcgtccccccc 797

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	740.4	92.9	887	14	BQ948158	BQ948158 AGENCOURT
2	656.4	82.4	1157	14	BQ940483	BQ940483 AGENCOURT
3	641.6	80.5	853	12	BG437538	BG437538 AGENCOURT
4	621	77.9	1053	12	BG164929	BG164929 AGENCOURT
5	590.8	74.1	594	14	BM763333	BM763333 K-EST0044
6	577.4	72.4	580	14	BM848189	BM848189 K-EST0127

7	571.4	71.7	573	14	BM840863	BM840863 K-EST0118
8	563	70.6	955	9	AL543945	AL543945 AL543945
9	528.8	66.3	532	14	BM846370	BM846370 K-EST0125
10	520.4	65.3	522	14	BM841897	BM841897 K-EST0119
11	518.4	65.0	542	14	BM821005	BM821005 K-EST0089
12	514.8	64.6	529	14	BM847924	BM847924 K-EST0127
13	458.8	57.6	569	14	BM846748	BM846748 K-EST0125
14	412	51.7	476	10	BE632644	BE632644 uv63f01.Y
15	405.4	50.9	420	14	BM764238	BM764238 K-EST0045
16	403.4	50.6	458	13	BM363136	BM363136 BS220054A
17	387	48.6	407	9	AI752561	AI752561 cni7d05.x
18	371.4	46.6	482	14	BM846622	BM846622 K-EST0125
19	277.6	34.8	913	13	BI912197	BI912197 603070588
20	274.4	34.4	512	9	AI390475	AI390475 mu59g08.Y
21	227.2	28.5	488	9	AA015243	AA015243 mh30c11.X
22	227	28.5	291	10	BB864730	BB864730 BB864730
23	226	28.4	915	12	BF035982	BF035982 601457860
24	174.4	21.9	552	12	BG095271	BG095271 uu85907.X
25	165.2	20.7	440	12	BG148676	BG148676 uu85907.Y
26	161	20.2	691	12	BF213570	BF213570 601845370
27	160.2	20.1	560	9	AL800340	AL800340 AL800340
28	116	14.6	655	13	BJ096761	BJ096761 BJ096761
29	114.6	14.4	633	13	BJ099801	BJ099801 BJ099801
30	63.4	8.0	711	12	BG758365	BG758365 602712557
31	57	7.2	925	17	CNS0091P	AL053013 Drosophil
32	56.8	7.1	180	9	AA204015	AA204015 mu59g08.X
33	52.2	6.5	925	17	CNS0091P	AL053013 Drosophil
34	50.2	6.3	1023	17	AG128304	AG128304 Pan trogl
35	49.6	6.2	816	17	CNS037NP	AL231550 Tetraodon
36	48.4	6.1	935	17	AG126234	AG126234 Pan trogl
37	48.4	6.1	1041	14	BQ058321	BQ058321 AGENCOURT
38	48.4	6.1	1469	13	BM552519	BM552519 AGENCOURT
39	48.2	6.0	1090	17	AG072433	AG072433 Pan trogl
40	48	6.0	958	9	AL547011	AL547011 AL547011
41	47.2	5.9	927	14	BQ680645	BQ680645 AGENCOURT
42	47.2	5.9	1034	14	BQ641825	BQ641825 AGENCOURT
43	47	5.9	1651	12	BG809816	BG809816 mgct001XX
44	46.8	5.9	941	14	BQ954342	BQ954342 AGENCOURT
45	46.8	5.9	955	14	BQ949322	BQ949322 AGENCOURT

## ALIGNMENTS

RESULT 1 BQ948158 887 bp mRNA linear EST 21-AUG-2002  
LOCUS AGENCOURT\_8813192 NIH\_MGC\_101 Homo sapiens cDNA clone IMAGE:6428214  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ948158  
VERSION BQ948158.1 GI:22363636  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 887)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing By: The I.M.A.G.E. Consortium (LLNL)  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLCM2614 row: h column: 07  
High quality sequence stop: 674.  
Location/Qualifiers  
1. 887

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6428214"
/tissue_lib="NIH MGC_101"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT 154 a 325 c 240 g 167 t 1 others
ORIGIN
Query Match 92.9%; Score 740.4; DB 14; Length 887;
Best Local Similarity 99.7%; Pred. No. 5.9e-151;
Matches 752; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 ATTAAAGCTTCGCGAGCGCGCTCGCCCTCCACTCCGACGCTCCGCGAGAGGAG 60
Db 8 ATTAAAGCTTCGCGAGCGCGCTCGCCCTCCACTCCGACGCTCCGCGAGAGGAG 67
QY 61 CCGCACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 68 CCGCACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 127
QY 121 TGTTAGCGTGTGTGACCGGTGTCTGGACCTCTCCGACCTCTCCGAGTGCAGCTCTCAATCGCA 180
Db 128 TGTTAGCGTGTGTGACCGGTGTCTGGACCTCTCCGACCTCTCCGAGTGCAGCTCTCAATCGCA 187
QY 181 CAGGGACCCAGGGCGCTGGCGCTCCATCCAGAAACCTATGACCTACCCGCTACCTGG 240
Db 188 CAGGGACCCAGGGCGCTGGCGCTCCATCCAGAAACCTATGACCTACCCGCTACCTGG 247
QY 241 AGCACCACCTCCGACGCTGGTGGAGCTATCTGAATACCTGGGCGCGCGCTTTCAACG 300
Db 248 AGCACCACCTCCGACGCTGGTGGAGCTATCTGAATACCTGGGCGCGCGCTTTCAACG 307
QY 301 AGCCAGACTTCAACCTCCCGCC--TGGGGGAGAGACTCTGCCAGGGCGCACTGTTGAC 359
Db 308 AGCCAGACTTCAACCTCCCGCCATGGGGGAGAGACTCTGCCAGGGCGCACTGTTGAC 367
QY 360 TTGGAGGTGTGGGAGAGCTCAATGACAAACTCGCGCTGACCCAGAACTACGAGGCTTAC 419
Db 368 TTGGAGGTGTGGGAGAGCTCAATGACAAACTCGCGCTGACCCAGAACTACGAGGCTTAC 427
QY 420 AGCCACCTTCTGTGTTACTTGGTGGCGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGC 479
Db 428 AGCCACCTTCTGTGTTACTTGGTGGCGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGC 487
QY 480 CGCAGCTGGCCCACTTCTGACACAGCTCCAGGGCGCTGTGGGCGAGCAATTGGGGCGTC 539
Db 488 CGCAGCTGGCCCACTTCTGACACAGCTCCAGGGCGCTGTGGGCGAGCAATTGGGGCGTC 547
QY 540 ATGGGAGCTCTGGGCTACCCACTGCCCGCAGCGCTGCTGGGACTGAACCCACTTGGACT 599
Db 548 ATGGGAGCTCTGGGCTACCCACTGCCCGCAGCGCTGCTGGGACTGAACCCACTTGGACT 607
QY 600 CTGGGCGCTGCCACAGTGAATTCCTCCAGAGTGAACGACTTCTGGCTGCTGAAGGAG 659
Db 608 CTGGGCGCTGCCACAGTGAATTCCTCCAGAGTGAACGACTTCTGGCTGCTGAAGGAG 667
QY 660 CTGCAGACTTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCGAG 719
Db 668 CTGCAGACTTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCGAG 727
QY 720 CTTCCAGAGCTGAGTCACTGCTGCACTGGGG 753
Db 728 CTTCCAGAGCTGAGTCACTGCTGCACTGGGG 761
```

```
RESULT 2
BQ940483
LOCUS
DEFINITION
AGENCOURT 8864294 Lupski_sciatic_nerve Homo sapiens cDNA clone
IMAGE:6197786 5', mRNA sequence.
ACCESSION
BQ940483
VERSION
BQ940483.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1157)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13608 row: g column: 03
High quality sequence stop: 572.
Location/Qualifiers
1..1157
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6197786"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGTCCG-3' and
5'-GACTAGTCTTAGATCGCGCGCGCGCTCTGCGACGCTGCTGCGACCTCCCTGCACT 163
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
```

```
BASE COUNT 221 a 374 c 346 g 216 t
ORIGIN
Query Match 82.4%; Score 656.4; DB 14; Length 1157;
Best Local Similarity 97.4%; Pred. No. 1.2e-132;
Matches 678; Conservative 0; Mismatches 16; Indels 2; Gaps 1;
QY 104 AGGGGACTCTGGGGGATGTTAGCGTGTGTCACGCTGCTGCGACCTCCCTGCACT 163
Db 58 AGGGGACTCTGGGGGATGTTAGCGTGTGTCACGCTGCTGCGACCTCCCTGCACT 117
QY 164 GCCAGCTCTCAATCGCAGGGGACCCAGGGCGCTGGCGCTCCATCCAGAAACCTATGA 223
Db 118 GCCAGCTCTCAATCGCAGGGGACCCAGGGCGCTGGCGCTCCATCCAGAAACCTATGA 177
QY 224 CCTCACCCTGCTTACGAGCACCACCTCGGAGCTTGGCTGGGACCTATCTGAACCTACCT 283
Db 178 CCTCACCCTGCTTACGAGCACCACCTCGGAGCTTGGCTGGGACCTATCTGAACCTACCT 237
QY 284 GGGCCCCCTTTCACAGGCGAGACTTCAACCTCCCGCTGGGGCGAGAGACTCTGCC 343
Db 238 GGGCCCCCTTTCACAGGCGAGACTTCAACCTCCCGCTGGGGCGAGAGACTCTGCC 297
QY 344 CAGGGCCTGTTGACTTGGAGGTGTGGGAGAGCTCAATGACAACTGCGGCTGACCA 403
```



Db 542 CTACCCACTGCCCCCAGCGCTGCTGGAGTGAACCCACTTTGGACTCTTGGCCCTTGC 601  
Qy 612 CACAGTCACTCTCTCCAGAGATGACGACCTCTGGCTGCTGAAGAGAGCTGCAGACCTGG 671  
Db 602 CACAGTCACTCTCTCCAGAGATGACGACCTCTGGCTGCTGAAGAGAGCTGCAGACCTGG 661  
Qy 672 CTGTGGCGCTGGCCCAAGGACTTCAACCGGCTCAAGAAAGAGATCGAGCTCCAGCAGCT 731  
Db 662 CTGTGGCGCTGGCC-AGGACTTCCACGGTCCAGACGAAGATTGCAGCTCCAGCAGTT 720  
Qy 732 GCAGTCACTCTGACCTGGCGCTGATGGCTTCTGACTTCTGACTT 778  
Db 721 GCAGTCACTCTGACCTGGCGCTGATGGCTTCTGACTTCTGACTT 766

RESULT 5  
BM763333  
LOCUS K-EST0044645 S13KMS5 Homo sapiens cDNA clone S13KMS5-18-E09 5',  
DEFINITION mRNA sequence.  
ACCESSION BM763333  
VERSION BM763333.1 GI:19092948  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 594)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished (2002)  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 18 row: E column: 09  
High quality sequence stop: 594.  
Location/Qualifiers

FEATURES  
source  
1..594  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="S13KMS5-18-E09"  
/clone\_lib="S13KMS5"  
/tissue\_type="myeloma"  
/cell\_line="KMS-5"  
/lab\_host="Top10F"  
/note="vector: pcNS; Site\_1: EcoRI; Site\_2: NotI; The poly  
(A) + RNA was dephosphorylated with bacterial alkaline  
phosphatase (BAP) and then deapped with tabacco acid  
pyrophosphatase (TAP). The deapped intact mRNA was  
ligated with DNA-RNA linker including EcoR I site by  
treatment of T4 RNA ligase and the first strand cDNA was  
synthesized from oligo dt-tailed mRNA by priming with  
dt-tailed vector. The dt-tailed vector was adjusted to  
have about 60nt. The cDNA vector of EcoRI which site is  
coli DNA ligase after digestion of EcoRI which site is  
also included in vector. An RNA strand converted to a DNA  
strand by Okayama-Berg method. The obtained cDNA vectors  
were used for transformation of competent cells E. coli  
Top10F, by electroporation method. The cDNA libraries  
constructed by this method are full-length enriched cDNA  
library."

BASE COUNT 101 a 226 c 164 g 103 t  
ORIGIN  
Query Match 74.1%; Score 590.8; DB 14; Length 594;  
Best Local Similarity 99.7%; Pred. No. 1.8e-118;

1 (bases 1 to 1053)  
NIH-MGC http://mgc.nci.nih.gov/  
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)  
UNPUBLISHED (1999)  
CONTACT: ROBERT STRAUSBERG, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLNL0244 row: m column: 14  
High quality sequence start: 3  
High quality sequence stop: 675.  
Location/Qualifiers  
1..1053  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_89"  
/tissue\_type="hypernephroma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: kidney; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dt primed.  
Average insert size 1.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
BASE COUNT 210 a 351 c 292 g 200 t  
ORIGIN

Query Match 77.9%; Score 621; DB 12; Length 1053;  
Best Local Similarity 94.1%; Pred. No. 5.9e-125;  
Matches 722; Conservative 0; Mismatches 35; Indels 10; Gaps 7;  
Qy 20 CGCGGCTCGCCCTCCATCCGAGCTCCGGAGAGAGCGGACCGCGCGGCCAG 79  
Db 2 CGCGGCTCGCCCTCCATCCGAGCTCCGGAGAGAGCGGACCGCGCGGCCAG 61  
Qy 80 CCCAG-CCCCATGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGGCTGCTGCA 138  
Db 62 CCCCAGTCCCCATGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGGCTGCTGCA 121  
Qy 139 CGGTGCTCTGGCACTCTCCGAGTCCGAGCTTCAATGCGACAGGGACCGAGCGCTG 198  
Db 122 CGGTGCTCTGGCACTCTCCGAGTCCGAGCTTCAATGCGACAGGGACCGAGCGCTG 181  
Qy 199 GCCCTCCATCCAGAAACCTATGACCTCACCGGCTACCTGGAGCACCACCTCCGAGCT 258  
Db 182 GCCCTCCATCCAGAAACCTATGACCTCACCGGCTACCTGGAGCACCACCTCCGAGCT 241  
Qy 259 TGGCTGGGACCTATCTGAATCTCTGGGCCCCCCCC---TTTCAAGCAGGACGACTTCAAC 314  
Db 242 TGGCTGGGACCTATCTGAATCTCTGGGCCCCCCCCTTTCAAGCAGGACTTCAAC 301  
Qy 315 CTTCCCGGCTGGGGCAGAGACTCTG-CCGAGGCGACTGTTGACTTGGAGGTGTGGCG 373  
Db 302 CTTCCCGGCTGGGGCAGAGACTCTGTTCCAGGGGCCACTGTTGACTTGGAGGTGTGGCG 361  
Qy 374 AGCCCTCAATGACAACTCGGGCTGACCCAGAACTACGAGCGCTTACAGCCACTTCTGTG 433  
Db 362 AGCCCTCAATGACAACTCGGGCTGACCCAGAACTACGAGCGCTTACAGCCACTTCTGTG 421  
Qy 434 TTAATTGCGTGGCTCAACCGTCAAGCTGCGACTGCTGAGTGGCGCGGAGCTGGCCCA 493  
Db 422 TTAATTGCGTGGCTCAACCGTCAAGCTGCGACTGCTGAGTGGCGCGGAGCTGGCCCA 481  
Qy 494 CTTCTGACACGAGCTCCAGGCGCTGCTGGGAGCATTCGCGGCGCTCATGGAGCTCTGGG 553  
Db 482 GTTCTGACACGAGCTCCAGGCGCTGCTGGGAGCATTCGCGGCGCTCATGGAGCTCTGGG 541  
Qy 554 CTACCCACTG-CCCCAGCGGCTGCTGGGACTGAACCCAC-TTGGACTCTCTGGCCCTGCC 611

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT



QY	Db	241	ATCTGAAC	TACCTGGGCC	CCCCCTTTCAA	CGAGCCAGAC	TTCAACCC	TCCCGCCTGGGG	300
b		331	CAGAGACT	CTGCGCCAGGGCCAC	TGTTGACTT	TGAGGTGTGGG	GAAGCCCTCAATGACA	AAAC	390
Y		301	CAGAGACT	CTGCGCCAGGGCCAC	TGTTGACTT	TGAGGTGTGGG	GAAGCCCTCAATGACA	AAAC	360
b		391	TGCGGTG	TACCCAGAACTAGAGG	CCCTACAGGCAC	CTTCTGTGTTACT	TTCGTTGGGCTCA	450	
Y		361	TGCGGTG	TACCCAGAACTAGAGG	CCCTACAGGCAC	CTTCTGTGTTACT	TTCGTTGGGCTCA	420	
b		451	ACGTCAGG	CTTGCCACCTGCTGAG	CTGCGCCGAC	CGCTTGTGAC	CCAGCCCTCC	510	
Y		421	ACGTCAGG	CTTGCCACCTGCTGAG	CTGCGCCGAC	CGCTTGTGAC	CCAGCCCTCC	480	
b		511	AGGCGCT	CTGCTGGG	CAGATTGCGGG	CGTCATGCGCAG	CTCTGGGCTACCC	ACTGCCCCAGC	570
Y		481	AGGCGCT	CTGCTGGG	CAGATTGCGGG	CGTCATGCGCAG	CTCTGGGCTACCC	ACTGCCCCAGC	540
b		571	CGTGCCT	TGGGACTGA	ACCCACTTTGG	ACTCTCTG	603		
Y		541	CGTGCCT	TGGGACTGA	ACCCACTTTGG	ACTCTCTG	573		
b									
RESULT 8		AL543945				955 bp	mRNA	linear	EST 16-FEB-2001
LOCUS		AL543945	171	NFL006	PL2	955	bp	mRNA	linear
DEFINITION		AL543945	171	NFL006	PL2	955	bp	mRNA	linear
ACCESSION		AL543945	171	NFL006	PL2	955	bp	mRNA	linear
VERSION		AL543945.1	171	NFL006	PL2	955	bp	mRNA	linear
KEYWORDS		EST.	171	NFL006	PL2	955	bp	mRNA	linear
SOURCE		human.	171	NFL006	PL2	955	bp	mRNA	linear
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	171	NFL006	PL2	955	bp	mRNA	linear
REFERENCE		1 (bases 1 to 955)	171	NFL006	PL2	955	bp	mRNA	linear
AUTHORS		Li, W.B., Gruber, C., Jessee, J. and Polayes, D.	171	NFL006	PL2	955	bp	mRNA	linear
TITLE		Full-length cDNA libraries and normalization	171	NFL006	PL2	955	bp	mRNA	linear
JOURNAL		Unpublished (2001)	171	NFL006	PL2	955	bp	mRNA	linear
COMMENT		Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.	171	NFL006	PL2	955	bp	mRNA	linear
FEATURES		Location/Qualifiers	171	NFL006	PL2	955	bp	mRNA	linear
source		1..955	171	NFL006	PL2	955	bp	mRNA	linear
		/organism="Homo sapiens"	171	NFL006	PL2	955	bp	mRNA	linear
		/db_xref="taxon:9606"	171	NFL006	PL2	955	bp	mRNA	linear
		/clone="CS0D1004YM15"	171	NFL006	PL2	955	bp	mRNA	linear
		/clone_lib="LTI_NFL006_PL2"	171	NFL006	PL2	955	bp	mRNA	linear
		/tissue_type="placenta"	171	NFL006	PL2	955	bp	mRNA	linear
		/notes="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"	171	NFL006	PL2	955	bp	mRNA	linear
BASE COUNT		168 a 323 c 266 g 198 t	171	NFL006	PL2	955	bp	mRNA	linear
ORIGIN			171	NFL006	PL2	955	bp	mRNA	linear
Query Match		70.6%;	171	NFL006	PL2	955	bp	mRNA	linear
Best Local Similarity		99.7%;	171	NFL006	PL2	955	bp	mRNA	linear
Matches		575; Conservative	171	NFL006	PL2	955	bp	mRNA	linear
0; Mismatches		0; Indels	171	NFL006	PL2	955	bp	mRNA	linear
2; Gaps			171	NFL006	PL2	955	bp	mRNA	linear
QY		219	TATGACCT	CACCCGCTACTG	GAGACCA	CTCCGAC	TTGGCTGGG	ACCTATCTGAAC	278
Db		1	TATGACCT	CACCCGCTACTG	GAGACCA	CTCCGAC	TTGGCTGGG	ACCTATCTGAAC	60
QY		279	TACCTGGG	CCCCCTTTCAAC	GACGAC	ACTTCAAC	CCCTCCCGCCT	TGGGGGACAGACT	338





LOCUS BM841897 522 bp mRNA linear EST 06-MAR-2002  
DEFINITION K-EST0119216 S13KMS5 Homo sapiens cDNA clone S13KMS5-34-A05 5',  
mRNA sequence.  
ACCESSION BM841897  
VERSION BM841897.1 GI:19198306  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 522)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished (2002)  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 34 row: A column: 05  
High quality sequence stop: 522.  
Location/Qualifiers  
1. .522  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S13KMS5-34-A05"  
/clone\_lib="S13KMS5"  
/tissue\_type="myeloma"  
/cell\_line="KMS-5"  
/lab\_host="Top10F"  
/note="vector: pCNS; Site 1: EcoRI; Site 2: NotI; The poly  
(A)+ RNA was dephosphorylated with bacterial alkaline  
phosphatase (BAP) and then decapped with tobacco acid  
pyrophosphatase (TAP). The decapped intact mRNA was  
ligated with DNA-RNA linker including EcoR I site by  
treatment of T4 RNA ligase and the first strand cDNA was  
synthesized from oligo dt-selected mRNA by priming with  
dt-tailed vector. The dt-tailed vector was circularized with E.  
have about 60nt. The cDNA vector was circularized with E.  
coli DNA ligase after digestion of EcoRI which site is  
also included in vector. An RNA strand converted to a DNA  
strand by Okayama-Berg method. The obtained cDNA vectors  
were used for transformation of competent cells E. coli  
Top10F' by electroporation method. The cDNA libraries  
constructed by this method are full-length enriched cDNA  
library."

271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCTGGGG 330  
241 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCTGGGG 300  
331 CAGAGACTCTGCCAGGGCCACTGTTGACTTTGGAGGTGGCGAAGCCTCAATGACAAAC 390  
301 CAGAGACTCTGCCAGGGCCACTGTTGACTTTGGAGGTGGCGAAGCCTCAATGACAAAC 360  
391 TGGCGCTGACCCAGAGAACTACGAGGCTACAGCCACTTCTGTGTACTTGGCTGCTCA 450  
361 TGGCGCTGACCCAGAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGGCTGCTCA 420  
451 ACCGTGAGCTGCGCACTGCTGAGCTGCGCGCAGCGCTGCCACCTTCTGCACAGCTCC 510  
421 ACCGTGAGCTGCGCACTGCTGAGCTGCGCGCAGCGCTGCCACCTTCTGCACAGCTCC 480  
511 AGGGCTGCTGGCGAGCATTTGCGGGCGTTCATGGAGCTCTGG 552  
481 AGGGCTGCTGGCGAGCATTTGCGGGCGTTCATGGAGCTCTGG 522  
RESULT 11  
BM821005 542 bp mRNA linear EST 06-MAR-2002  
LOCUS K-EST0089713 S19N665307 Homo sapiens cDNA clone S19N665307-9-G05  
DEFINITION 5', mRNA sequence.  
ACCESSION BM821005  
VERSION BM821005.1 GI:19177418  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 542)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished (2002)  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 9 row: G column: 05  
High quality sequence stop: 542.  
Location/Qualifiers  
1. .542  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S19N665307-9-G05"  
/clone\_lib="S19N665307"  
/sex="W"  
/lab\_host="Top10F"  
/note="Organ: Stomach; Vector: pCNS; Site 1: EcoRI;  
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tobacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including EcoR  
I site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized from oligo dt-selected mRNA by  
priming with dt-tailed vector. The cDNA vector was  
adjusted to have about 60nt. The cDNA vector was  
circularized with E. coli DNA ligase after digestion of  
EcoRI which site is also included in vector. An RNA strand  
converted to a DNA strand by Okayama-Berg method. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10F' by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."

BASE COUNT 92 a 202 c 155 g 92 t



K-EST0125824 S13KMS5 Homo sapiens cDNA clone S13KMS5-40-G02 5',  
 mRNA sequence.  
 BM846748.1 GI:19203147  
 EST.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 569)  
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,  
 Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and  
 Kim, Y.S.  
 21C Frontier Korean EST Project 2001  
 Unpublished (2002)  
 Contact: Kim YS  
 Genome Research Center  
 Korea Research Institute of Bioscience & Biotechnology  
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
 Tel: +82-42-860-4470  
 Fax: +82-42-860-4409  
 Email: yongsung@mail.krrib.re.kr  
 Plate: 40 row: G column: 02  
 High quality sequence stop: 569.  
 Location/Qualifiers  
 1..569  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="S13KMS5-40-G02"  
 /clone\_lib="S13KMS5"  
 /tissue\_type="myeloma"  
 /cell\_line="KMS-5"  
 /lab\_host="Top10p;"  
 /note="vector: PCNS; Site\_1: EcoRI; Site\_2: NotI; The poly  
 (A)+ RNA was dephosphorylated with bacterial alkaline  
 phosphatase (BAP) and then decapped with tobacco acid  
 pyrophosphatase (TAP). The decapped intact mRNA was  
 ligated with DNA-RNA linker including EcoRI site by  
 treatment of T4 RNA ligase and the first strand cDNA was  
 synthesized from oligo dT-selected mRNA by priming with  
 dT-tailed vector. The dT-tailed vector was circularized with E.  
 coli DNA ligase after digestion of EcoRI which site is  
 also included in vector. An RNA strand converted to a DNA  
 strand by Okayama-Berg method. The obtained cDNA vectors  
 were used for transformation of competent cells E. coli  
 'Top10p' by electroporation method. The cDNA libraries  
 constructed by this method are full-length enriched cDNA  
 library."  
 BASE COUNT 135 a 192 g 151 g 91 t  
 ORIGIN  
 Query Match 57.6%; Score 458.8; DB 14; Length 569;  
 Best Local Similarity 99.6%; Pred. No. 9.4e-90;  
 Matches 460; Conservative 0; Mismatches 2; Indels 0; Gaps 0  
 Oy 1 ATTTAAAGCTTCGCGAGCGCGCTGCCCTCCCACTCCGCAGAGCCTCGGAGAGGAG 60  
 Db 108 ATTTAAAGCTTCGCGAGCGCGCTGCCCTCCCACTCCGCAGAGCCTCGGAGAGGAG 167  
 Oy 61 CCGCACCCGGCGGCCCCAGAGCCCACTGGACCTCCGAGCAGGGGAGCTCGTGGGGGA 120  
 Db 168 CCGCACCCGGCGGCCCCAGAGCCCACTGGACCTCCGAGCAGGGGAGCTCGTGGGGGA 227  
 Oy 121 TGTATTAGCGTGCTGTGACAGCGTGCTCTGGACCTCTCCCTGCAGTGCCAGCTCAATGCCA 180  
 Db 228 TGTATTAGCGTGCTGTGACAGCGTGCTCTGGACCTCTCCCTGCAGTGCCAGCTCTCAATGCCA 287  
 Oy 181 CAGGGGACCCAGGCGCTTGGCCCTCCATCCAGAAAACTATGACCTCACCGCTACTCGT 240  
 Db 288 CAGGGGACCCAGGCGCTTGGCCCTCCATCCAGAAAACTATGACCTCACCGCTACTCGT 347  
 Oy 241 AGCAACAACTCCGACAGCTTGGCTGGGACCTATCTGAACACTCTGGGCCCCCTTTCAACG 300





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 202.966 Seconds  
(without alignments)  
8043.084 Million cell updates/sec

Title: US-09-931-704-1  
Perfect score: 797  
Sequence: 1 attaaagcttcgcggagcc.....tctctcttcgctccccccc 797

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID82/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID82/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID82/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID82/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID82/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID82/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID82/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID82/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID82/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID82/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID82/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID82/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID82/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID82/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID82/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID82/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID82/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID82/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID82/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID82/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID82/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID82/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID82/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID82/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	797	100.0	797	19 AAV47510	Human neurotrophic
2	797	100.0	797	19 AAV22652	CDNA encoding huma
3	797	100.0	797	21 AAA39481	Human NNT-1 CDNA.
4	797	100.0	797	24 ABK11647	Human CDNA encodin
5	793.4	99.5	881	22 AAH74484	Nucleotide sequenc
6	793.4	99.5	1790	21 AAA88546	Human interleukin-
7	759.2	95.3	768	22 AAH99772	Human protein enco
8	749.4	94.0	1710	20 AAX16161	Human cardiotoxop
9	710	89.1	729	22 AAD04201	Human cardiotoxop

10	694.8	87.2	1008	22	AAKS1548	Human polynucleoti
11	669.4	84.0	819	19	AAV47512	Mouse neurotrophic
12	669.4	84.0	819	19	AAV22654	CDNA encoding muri
13	669.4	84.0	819	21	AAA39483	Murine NNT-1 CDNA.
14	669.4	84.0	819	24	ABK11649	Mouse CDNA encodin
15	668.4	83.9	968	22	ABA09140	Human cardiotoxop
16	668.4	83.9	968	22	AAKS2532	Human polynucleoti
17	558.4	70.1	648	21	AAA88547	Mouse interleukin-
18	523.4	65.7	5087	21	AAA39482	Human NNT-1 DNA.
19	523.4	65.7	5087	24	ABK11648	Human novel neurot
20	523.4	65.7	5088	19	AAV47511	Human neurotrophic
21	523.4	65.7	5088	19	AAV22653	Human genomic DNA
22	493.4	61.9	495	22	ABA17773	Human foetal liver
23	493.4	61.9	495	22	ABA37855	Probe #16321 for g
24	493.4	61.9	495	22	AAK20142	Human brain expres
25	493.4	61.9	495	22	AAK46202	Human bone marrow
26	493.4	61.9	495	22	AAI25564	Probe #15497 for g
27	493.4	61.9	495	22	AAI52108	Probe #20794 used
28	493.4	61.9	495	24	ABS20513	Human genome-deriv
29	369	46.3	492	22	ABA59256	Human foetal liver
30	369	46.3	492	22	ABA27996	Human brain expres
31	369	46.3	492	22	AAK07469	Human brain expres
32	369	46.3	492	22	AAK33253	Human bone marrow
33	369	46.3	492	22	AAI16384	Probe #8317 for ge
34	369	46.3	492	22	AAI39047	Probe #7733 used t
35	369	46.3	492	24	ABS08085	Human genome-deriv
36	269.4	33.8	283	22	AAH23145	Osteoarthritis tis
37	80.2	10.1	396	20	AAX16162	Human cardiotoxop
38	73.6	9.2	1260	24	ABQ16902	Oligonucleotide fo
39	73.6	9.2	1260	24	ABQ16903	Oligonucleotide fo
40	58.6	7.4	65	24	ABNS6698	Mouse spliced tran
41	49	6.1	627	20	ABQ60797	Human DNAX interle
42	46.8	5.9	1260	24	ABQ16904	Oligonucleotide fo
43	46.8	5.9	1260	24	ABQ16905	Oligonucleotide fo
44	46.8	5.9	2320	21	AAE21879	Human breast and o
45	46.2	5.8	627	20	AAE60796	Human DNAX interle

## ALIGNMENTS

RESULT 1  
AAV47510  
ID AAV47510 standard; cDNA; 797 BP.

AC AAV47510;

XX 09-NOV-1998 (first entry)

DE Human neurotrophic factor NNT-1 CDNA.

XX NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
KW peripheral neuropathy; dystrophy; neural retina degeneration;  
KW common variable immunodeficiency; CVID; selective IgA deficiency;  
KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;  
KW therapy; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
FH CDS 90...767

FT sig\_peptide /\*tag= a

FT mat\_peptide /\*tag= b

FT WO9833922-A1.

PD 06-AUG-1998.

XX





XX Disclosure; Fig 1; 41pp; English.

XX The present sequence encodes a human neurotrophic factor, designated

CC NNT-1, which is capable of stimulating growth of motor or sympathetic

CC neurons. The NNT-1 protein is useful in the treatment of neurological

CC diseases characterized by the degeneration and death of particular

CC classes of neurons. These diseases specifically include Parkinson's

CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,

CC stroke and various degenerative disorders affecting vision.

XX

XX Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;

XX

Query Match 100.0%; Score 797; DB 19; Length 797;

Best Local Similarity 100.0%; Pred. No. 1.1e-176;

Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACATCCGCGAGCTCCGCGGAGAGAG 60

DB 1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACATCCGCGAGCTCCGCGGAGAGAG 60

QY 61 CCGCACCGCGCGCGCCAGCCCGCCATGGACCTCCGAGCAGGGGACTCGTGGGGA 120

DB 61 CCGCACCGCGCGCGCCAGCCCGCCATGGACCTCCGAGCAGGGGACTCGTGGGGA 120

QY 121 TGTAGCGTGCCTGTGACAGGTGCTCTGGCACCTCCCTGCGAGTGCAGCTTCAATCGCA 180

DB 121 TGTAGCGTGCCTGTGACAGGTGCTCTGGCACCTCCCTGCGAGTGCAGCTTCAATCGCA 180

QY 181 CAGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240

DB 181 CAGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240

QY 241 AGCACCAACTCCCGAGCTTGGTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAACG 300

DB 241 AGCACCAACTCCCGAGCTTGGTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAACG 300

QY 301 AGCCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGCTGCTGCTGACT 360

DB 301 AGCCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGCTGCTGCTGACT 360

QY 361 TGGAGTGTGGCAAGCTCAATGAACAACTGGGGCTGAGCCAGAACTACGAGGCTTACA 420

DB 361 TGGAGTGTGGCAAGCTCAATGAACAACTGGGGCTGAGCCAGAACTACGAGGCTTACA 420

QY 421 GCCACCTTCTGTACTTGTGCGGCTCAACCGTCAGGCTGCTGCTGCTGCTGCTGCTG 480

DB 421 GCCACCTTCTGTACTTGTGCGGCTCAACCGTCAGGCTGCTGCTGCTGCTGCTGCTG 480

QY 481 GCAGCTGGCCCACTTCTGACAGGCTCCAGGCTGCTGCGGCTGCTGCGGCTGCTGCA 540

DB 481 GCAGCTGGCCCACTTCTGACAGGCTCCAGGCTGCTGCGGCTGCTGCGGCTGCTGCA 540

QY 541 TGGGAGCTTGGGCTACCACTGCCCCAGCGCTGCTGCGGCTGCTGCGGCTGCTGCACTC 600

DB 541 TGGGAGCTTGGGCTACCACTGCCCCAGCGCTGCTGCGGCTGCTGCGGCTGCTGCACTC 600

QY 601 TTGGCCCTTGGCCAGTGAATCTTCCAGAGATGAGGCTGCTGCGGCTGCTGCGGCTGCT 660

DB 601 TTGGCCCTTGGCCAGTGAATCTTCCAGAGATGAGGCTGCTGCGGCTGCTGCGGCTGCT 660

QY 661 TGCAGACTGGTGTGGGCTCGGCAAGGACTTCAACCGGCTCAAGAGAGATGCTGAGC 720

DB 661 TGCAGACTGGTGTGGGCTCGGCAAGGACTTCAACCGGCTCAAGAGAGATGCTGAGC 720

QY 721 CTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780

DB 721 CTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780

QY 781 CCTCTTCT 797

DB 781 CCTCTTCT 797

## RESULT 3

AAA39481

ID AAA39481 standard; cDNA; 797 BP.

XX AC AAA39481;

XX DT 24-AUG-2000 (first entry)

XX DE Human NNT-1 cDNA.

XX NNT-1; human; neurotrophic factor; neurotrophic; neuroprotective; treatment;

XX KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;

XX KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;

XX KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;

XX KW Huntington's disease; peripheral neuropathy; neural retina degeneration;

XX KW retinopathy; immune disorder; hematopoietic disorder; ss.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers

XX FT 90..767

XX FT /\*tag= a

XX FT /product= "NNT-1"

XX US6054294-A.

XX PD 25-APR-2000.

XX PF 12-DEC-1997; 97US-0988819.

XX PR 03-FEB-1997; 97US-0792019.

XX PA (AMGE-) AMGEN INC.

XX PI Chang M;

XX WP1; 2000-338492/29.

XX DR P-PSDB; AAY87813.

XX PT New nucleic acids encoding neurotrophic factors useful for stimulating

XX PT Growth of motor or sympathetic neurons for treating neuron cell damage

XX PS Claim 1a; Fig 1; 42pp; English.

XX CC This invention describes a novel nucleic acid molecule (I) encoding a

XX CC novel neurotrophic factor (NNT-1) (ii) which has neurotrophic,

XX CC neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and

XX CC ophthalmological activity. (I) is useful for producing NNT-1

XX CC polypeptides which are useful for treating patients in whom various

XX CC cells of the central, autonomic, or peripheral nervous system have

XX CC degenerated and/or have been damaged by congenital disease, trauma,

XX CC mechanical damage, surgery, stroke, ischemia, infection, metabolic

XX CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1

XX CC proteins are used to treat diseases like Alzheimer's, Parkinson's,

XX CC amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's

XX CC disease, peripheral neuropathy induced by diabetes or other metabolic

XX CC disorders, and/or dystrophies or degeneration of the neural retina such

XX CC as retinitis pigmentosa, drug-induced retinopathies, immune disorders and

XX CC night blindness, progressive cone-rod degeneration, stationary forms of

XX CC hematoepoietic disorders. (I) is effective in treating neurological

XX CC conditions and promotes neuron regeneration. Neural functions are

XX CC effectively restored in patients suffering from various neurological

XX CC disorders. This sequence encodes the human NNT-1 protein described in the

XX CC method of the invention.

XX SQ Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;

XX Query Match 100.0%; Score 797; DB 21; Length 797;

XX Best Local Similarity 100.0%; Pred. No. 1.1e-176;

XX Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

inflammatory bowel disease; transplant rejection; reproductive disorder; graft versus host disease; infertility; miscarriage; preterm labour.

Homo sapiens.

Key Location/Qualifiers  
90..767  
/\*tag= a  
/product= "NNT1"

WO200215977-A2.

28-FEB-2002.

17-AUG-2001; 2001WO-US25906.

18-AUG-2000; 2000US-226436P.

16-AUG-2001; 2001US-0931704.

(AMGE-) AMGEN INC.

Senaldi G;

WPI; 2002-280867/32.

P-PSDB; AAU78176.

Treating Immunoglobulin E-related disease, modulating IgE levels in a patient, preventing IgE-related disease and treating allergic diseases, involves administering NNT-1 inhibitor to a patient

Claim 2; Fig 1; 63pp; English.

The invention relates to treating Immunoglobulin E (IgE)-related disease, modulating IgE levels in a patient, preventing an IgE-related disease, and treating allergic diseases, comprising administering a therapeutically effective amount of novel neurotrophic factor (NNT)-1 inhibitor to a patient. Also included are a method of diagnosing an IgE-related disease or susceptibility to an IgE-related disease, by determining the presence or amount of expression of an NNT1 polypeptide encoded by a NNT1 nucleotide sequence, its fragment or naturally occurring variant, and diagnosing an IgE-related disease or susceptibility of an IgE-related disease based on the presence or amount of expression of the polypeptide and a pharmaceutical composition for use in treating IgE-related disease, comprising the NNT1 inhibitor.

The NNT1 inhibitor is useful for preventing and treating IgE-related disease, modulating IgE levels, and treating allergic diseases e.g. Type I allergic disease, allergic rhinitis, eczema, dermatitis, pollinosis, asthma, immune diseases and disorders, diseases involving abnormal cell proliferation including cancer, arteriosclerosis and vascular stenosis, diseases and conditions relating to dysfunction of immune system including rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, osteoarthritis, inflammatory joint disease, autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory bowel disease, transplant rejection, and graft versus host disease, and reproductive diseases and disorders including infertility, miscarriage, preterm labour and delivery, and endometriosis. The present sequence encodes human NNT1.

Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;

Query Match 100.0%; Score 797; DB 24; Length 797;  
Best Local Similarity 100.0%; Pred. No. 1.1e-176;  
Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCACTCCGACGCTCCGGAGAGGAG 60  
DB 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCACTCCGACGCTCCGGAGAGGAG 60  
QY 61 CGCAGCCCGCGCGCGCGGCTCGCCCTCCACTCCGACGCTCCGGAGAGGAG 120  
DB 61 CGCAGCCCGCGCGCGCGGCTCGCCCTCCACTCCGACGCTCCGGAGAGGAG 120  
QY 121 TGTAGCGTGCCTGTGACGCGGTCTCTGGCACTCCCTGTCAGTCCAGCTCTCAATCGCA 180

QY 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCACTCCGACGCTCCGGAGAGGAG 60  
DB 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCACTCCGACGCTCCGGAGAGGAG 60  
QY 61 CGCAGCCCGCGCGCGCGGCTCGCCCTCCACTCCGACGCTCCGGAGAGGAG 120  
DB 61 CGCAGCCCGCGCGCGCGGCTCGCCCTCCACTCCGACGCTCCGGAGAGGAG 120  
QY 121 TGTAGCGTGCCTGTGACGCGGTCTTGGCACTCCCTGAGTCCAGCTCTCAATCGCA 180  
DB 121 TGTAGCGTGCCTGTGACGCGGTCTTGGCACTCCCTGAGTCCAGCTCTCAATCGCA 180  
QY 181 CAGGGACCCAGCGGCTCGGCGGCTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240  
DB 181 CAGGGACCCAGCGGCTCGGCGGCTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240  
QY 241 AGCACAACCTCCGAGCTTGGCTGGGACCTATCTGAATACCTGGGCGGCTTCAACG 300  
DB 241 AGCACAACCTCCGAGCTTGGCTGGGACCTATCTGAATACCTGGGCGGCTTCAACG 300  
QY 301 AGCAGACTTCAACCTCCCGGCTGGGGGAGAGACTCTGCCAGGGGCCACTGTTGACT 360  
DB 301 AGCAGACTTCAACCTCCCGGCTGGGGGAGAGACTCTGCCAGGGGCCACTGTTGACT 360  
QY 361 TGGAGGTGTGGGAGGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCTTACA 420  
DB 361 TGGAGGTGTGGGAGGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCTTACA 420  
QY 421 GCCACCTTCTGTACTTGTGCTGCGCTCAACCGTCAGCTGCCACTGCTGAGCTGCGCC 480  
DB 421 GCCACCTTCTGTACTTGTGCTGCGCTCAACCGTCAGCTGCCACTGCTGAGCTGCGCC 480  
QY 481 GCAGCTGGGCGGCTTGTGACACGCTCCAGGGGCTGTGGGAGCAATGCGGGGCTCA 540  
DB 481 GCAGCTGGGCGGCTTGTGACACGCTCCAGGGGCTGTGGGAGCAATGCGGGGCTCA 540  
QY 541 TGGCAGCTTGGGCTACCACTGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 600  
DB 541 TGGCAGCTTGGGCTACCACTGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 600  
QY 601 CTGGCGCTGCCACAGTACTTCTCCAGAAAGATGACGACTTCTGGCTGCTGAAGGAGC 660  
DB 601 CTGGCGCTGCCACAGTACTTCTCCAGAAAGATGACGACTTCTGGCTGCTGAAGGAGC 660  
QY 661 TGCAGACTTGGCTGTGGCGCTCGGCCAGGACTTCAACCGGCTCAAGAGAAATGAGC 720  
DB 661 TGCAGACTTGGCTGTGGCGCTCGGCCAGGACTTCAACCGGCTCAAGAGAAATGAGC 720  
QY 721 CTCAGCAGCTGAGTCAACCTGACCTGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780  
DB 721 CTCAGCAGCTGAGTCAACCTGACCTGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780  
QY 781 CCTCTTCTGCTCCCGCCC 797  
DB 781 CCTCTTCTGCTCCCGCCC 797

RESULT 4

ID ABK11647 standard; cDNA; 797 BP.

XX ABK11647;

AC ABK11647;

05-JUN-2002 (first entry)

Human cDNA encoding novel neurotrophic factor NNT1.  
Human; ss; gene; NNT1; neurotrophic factor; IgE-related disease; Type I allergic disease; allergic rhinitis; eczema; dermatitis; pollinosis; asthma; immune disease; cancer; arteriosclerosis; vascular stenosis; rheumatoid arthritis; psoriatic arthritis; inflammatory arthritis; osteoarthritis; inflammatory joint disease; autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;

Db 121 TTTTAGCGTGTGTGACGGTGTCTGCGACCTCTGCGACCTCTGCGACCTCTCAATCGCA 180  
 Qy 181 CAGGGGACCCAGGGGCTGGCCCTTCCATCCAGAAACCTATGACCTCAACCTGCTACCTGG 240  
 Db 181 CAGGGGACCCAGGGGCTGGCCCTTCCATCCAGAAACCTATGACCTCAACCTGCTACCTGG 240  
 Qy 241 AGCACAACCTCGAGCTGGTGGGAGCTATCTGAACCTACCTGGGCCCCCTTTCAAG 300  
 Db 241 AGCACAACCTCGAGCTGGTGGGAGCTATCTGAACCTACCTGGGCCCCCTTTCAAG 300  
 Qy 301 AGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGACT 360  
 Db 301 AGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGACT 360  
 Qy 361 TGGAGGTGTGGGAGCTCAATGACAACTCGGCTGACCCAGAACTACGAGGCGCTACA 420  
 Db 361 TGGAGGTGTGGGAGCTCAATGACAACTCGGCTGACCCAGAACTACGAGGCGCTACA 420  
 Qy 421 GCCACCTTCTGTGTACTTGGTGGCTCAACCGTCAAGCTGCGACTGCTGAGCTGGGCC 480  
 Db 421 GCCACCTTCTGTGTACTTGGTGGCTCAACCGTCAAGCTGCGACTGCTGAGCTGGGCC 480  
 Qy 481 GCAGCTGGGCCACTTCTGACCAAGCTTCCAGGCGCTGCTGGGAGCATTCGGGGCTCA 540  
 Db 481 GCAGCTGGGCCACTTCTGACCAAGCTTCCAGGCGCTGCTGGGAGCATTCGGGGCTCA 540  
 Qy 541 TGGCAGCTTGGGCTACCACTGCGCCAGCGCTGCTGGGACTGAACCACTTGGACTC 600  
 Db 541 TGGCAGCTTGGGCTACCACTGCGCCAGCGCTGCTGGGACTGAACCACTTGGACTC 600  
 Qy 601 CTGGCCCTGGCCACAGTACTTCTCCAGAGATGGAGACTTCTGGCTGCTGAAGAGC 660  
 Db 601 CTGGCCCTGGCCACAGTACTTCTCCAGAGATGGAGACTTCTGGCTGCTGAAGAGC 660  
 Qy 661 TGCAGACTGGCTGTGGCTGCGCCAGAGACTTCAACCGGCTCAAGAAAGATGACG 720  
 Db 661 TGCAGACTGGCTGTGGCTGCGCCAGAGACTTCAACCGGCTCAAGAAAGATGACG 720  
 Qy 721 CTCACAGCTGGCTGTGGCTGCGCCAGAGACTTCAACCGGCTCAAGAAAGATGACG 780  
 Db 721 CTCACAGCTGGCTGTGGCTGCGCCAGAGACTTCAACCGGCTCAAGAAAGATGACG 780  
 Qy 781 CTCTTGGCTGCCCCC 797  
 Db 781 CTCTTGGCTGCCCCC 797

## RESULT 5

AAH74484

ID AAH74484 standard; DNA; 881 BP.

AC AAH74484;

 XX 15-OCT-2001 (first entry)  
 DT  
 DE

Nucleotide sequence of a human NNT-1 protein.

 KW NNT-1; CLF-1; SCNTFRalpha; nervous system; neuron; nervous system;  
 KW neuro-muscular function; tumour; immune system; haematopoietic system;  
 KW reproductive system; liver; skeletal muscle; neurodegenerative disease;  
 KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;  
 KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;  
 KW blastocyst implantation; thrombosis; retinal disease;  
 KW retinal pigmentosis; ss.

Homo sapiens.

 OS  
 XX  
 XX  
 Key Location/Qualifiers  
 CDS 174..851  
 FT /\*tag= a  
 FT /product= "NNT-1"  
 FT  
 XX

PN WO200155172-A2.  
 XX 02-AUG-2001.  
 PD 26-JAN-2001; 2001WO-FR00253.  
 PF 27-JAN-2000; 2000FR-0001035.  
 PR 12-OCT-2000; 2000FR-0013089.  
 XX (FABR) FABRE MEDICAMENT SA PIERRE.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;  
 XX WPT; 2001-488773/53.  
 DR P-PSDB; AAG63543.  
 XX  
 A complex comprising a NNT-1 protein and a CLF-1 and/or SCNTFRalpha  
 protein useful to treat neurodegenerative disease including Parkinson's  
 and Huntington's, obesity and cancer -  
 XX Disclosure; Page 57-58; 67pp; French.

The present sequence encodes a human NNT-1 protein. The specification  
 describes a complex comprising a NNT-1 protein and a CLF-1 and/or  
 SCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate  
 activity of the SCNTFRalpha/gp130/LiFrbeta receptor complex, or to  
 induce phosphorylation of the tyrosine of gp130 and LiFrbeta, or to  
 particularly where cells expressing the receptor complex are in the  
 central or peripheral nervous system, in neurons implicated in  
 neuro-muscular function or in skeletal muscle. The complex or  
 antibodies are also used to decrease the survival, growth or  
 proliferation of tumour cells or to facilitate the proliferation and/or  
 inhibit differentiation of cells stocks. The complex is also used to  
 modulate activity of the gp130/LiFrbeta receptor or cells expressing  
 that receptor, particularly those cells implicated in the immune,  
 haematopoietic, nervous or reproductive system, the liver or skeletal  
 muscle. Molecules of the invention may be used to prevent or treat  
 neurodegenerative diseases including amyotrophic lateral sclerosis,  
 Parkinson's and Huntington's disease, to repair or regenerate nervous  
 or muscular tissue or to maintain muscular mass in paralytic patients.  
 They may also be used to treat cancer, obesity and associated diseases,  
 and to improve fertility, particularly to avoid endometriosis and/or  
 assist blastocyst implantation, thrombosis, or retinal disease,  
 CC particular retinal pigmentosis.

SQ Sequence 881 BP; 158 A; 318 C; 246 G; 159 T; 0 other;

## Query Match

Best Local Similarity 99.5%; Score 793.4; DB 22; Length 881;

Matches 794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTTAAGCTTCGCGAGCGCGGCTCGCCCTCCACTCCGACCTCCGAGAGGAG 60  
 Db 85 ATTTAAGCTTCGCGAGCGCGGCTCGCCCTCCACTCCGAGAGGAG 144  
 Qy 61 CCGCACCCGCGCGCCAGCCCATGAGCTCCGAGCAGGAGACTCGTGGGGA 120  
 Db 145 CCGCACCCGCGCGCGCCAGCCCATGAGCTCCGAGCAGGAGACTCGTGGGGA 204  
 Qy 121 TGTTAGCTGCTGTGACAGGTGTCTGGACCTCCCTGACAGTGCAGCTCTCAATCGCA 180  
 Db 205 TGTTAGCTGCTGTGACAGGTGTCTGGACCTCCCTGACAGTGCAGCTCTCAATCGCA 264  
 Qy 181 CAGGGAGCCAGGCGCTGGCCCCCTCCATCCAGAAACCTATGACCTACCGCTACCTGG 240  
 Db 265 CAGGGAGCCAGGCGCTGGCCCCCTCCATCCAGAAACCTATGACCTACCGCTACCTGG 324  
 Qy 241 AGCAACCACTCCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAAG 300  
 Db 325 AGCAACCACTCCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAAG 384  
 Qy 301 AGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCGCCAGGCGCACTGTTGACT 360

Sun Feb 2 08:31:37 2003

us-09-931-704-1.rng

Db 385 AGCCAGACTTCAACCTCCGCCCTGGGGGACAGACTCTGCCAGGGCCACTGTTGACT 444  
Qy 361 TGGAGGTGGGGAAGCCCTCAATGACAAACTGGGGTGAAGCCAGAACTAGAGGCTTACA 420  
Db 445 TGGAGGTGGGGAAGCCCTCAATGACAAACTGGGGTGAAGCCAGAACTAGAGGCTTACA 504  
Qy 421 GCCACCTTCTGTGTTACTTGGCTGCTCAACCGTCAGGCTGCCACTGTGAGCTGGGCC 480  
Db 505 GCCACCTTCTGTGTTACTTGGCTGCTCAACCGTCAGGCTGCCACTGTGAGCTGGGCC 564  
Qy 481 GCAGCTGGCCCACTTCTGACACAGCTCCAGGGCTGCTGGGAGCAATTCGGGGCTTCA 540  
Db 565 GCAGCTGGCCCACTTCTGACACAGCTCCAGGGCTGCTGGGAGCAATTCGGGGCTTCA 624  
Qy 541 TGGCAGCTTGGGCTACCCACTGCCCGAGCGCTGCCCTGGGACTGAACCCACTTGGACTC 600  
Db 625 TGGCAGCTTGGGCTACCCACTGCCCGAGCGCTGCCCTGGGACTGAACCCACTTGGACTC 684  
Qy 601 CTGGCCCTGCCACAGTACTTCTCCAGAGATGGAGCACTTCTGGCTGCTGAAGGAGC 660  
Db 685 CTGGCCCTGCCACAGTACTTCTCCAGAGATGGAGCACTTCTGGCTGCTGAAGGAGC 744  
Qy 661 TGCAGACTTGGCTGGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAAAGATGCGAGC 720  
Db 745 TGCAGACTTGGCTGGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAAAGATGCGAGC 804  
Qy 721 CTCAGAGCTGACAGTACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACTTCT 780  
Db 805 CTCAGAGCTGACAGTACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACTTCT 864  
Qy 781 CCTCTGCTCCCC 795  
Db 865 CCTCTGCTCCCC 879

RESULT 6

AA88546  
ID AA88546 standard; DNA; 1790 BP.  
XX AC  
XX AA88546;  
XX 22-JAN-2001 (first entry)  
XX Human interleukin-B60 (IL-B60) gene.  
XX DE  
XX Interleukin-B60; IL-B60; human; cytokine; chromosome 11;  
KW cytokine-like factor-1; haematopoietic; inflammation;  
KW antiinflammatory; autoimmune disease; therapy; ds.  
XX OS  
XX Homo sapiens.  
XX FH Key Location/Qualifiers  
XX CDS 162..809  
XX FT /\*tag= a  
XX FT sig\_peptide 162..212  
XX FT /\*tag= b  
XX FT mat\_peptide 213..806  
XX FT /\*tag= c  
XX XX WO200053631-A1.  
XX PN  
XX 14-SEP-2000.  
XX PD  
XX 09-MAR-2000; 2000WO-US06182.  
XX PF  
XX 11-MAR-1999; 99US-0267901.  
XX PR  
XX (SCHE ) SCHERING CORP.  
XX PA Oppmann B, Timans JC, Kastelein RA, Bazan JF;  
XX PI WPI; 2000-587426/55.  
XX XX

DR P-PSDB; AAB19686.  
XX Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes,  
PT polypeptides, and nucleic acids, useful in research, diagnosis and for  
PT treating inflammatory and autoimmune disorders -  
XX  
XX Claim 17; Page 15-16; 97pp; English.  
PS

XX The present sequence is that of DNA encoding human interleukin-B60  
CC (IL-B60, see AAB19586), a novel, small soluble cytokine-like protein  
CC of 198 amino acids that exhibits structural motifs characteristic  
CC of a member of the long-chain cytokines, and which shows homology  
CC to granulocyte colony stimulating factor and interleukin-6. IL-B60  
CC may have either stimulatory or inhibitory effects on haematopoietic  
CC cells, including e.g. lymphoid cells, such as T-cells, B-cells,  
CC natural killer cells, macrophages, dendritic cells, haematopoietic  
CC progenitors, etc. Methods are provided for modulating the  
CC physiology or development of a cell or tissue culture cells by  
CC contacting the cell with an agonist or antagonist of IL-B60 or an  
CC agonist of antagonist of a complex of mature IL-B60 and its  
CC partner, cytokine-like factor-1 (CLF-1, see AAB19588). The  
CC IL-B60/CLF-1 cytokine serves as a key physiological factor in motor  
CC neuron development and regeneration. IL-B60, its agonists and  
CC antagonists may be used to treat inflammatory or autoimmune  
CC disorders and also for drug screening. The IL60B gene maps to  
CC human chromosome 11.  
XX

XX Sequence 1790 BP; 381 A; 560 C; 474 G; 375 T; 0 other;

Query Match 99.5%; Score 793.4; DB 21; Length 1790;  
Best Local Similarity 99.9%; Pred. No. 9.1e-176;  
Matches 794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTAAGCTTCGGCGGAGCGCGCTCGCCCTCCACTCCGCCAGCTCCGGGAGAGGAG 60  
Db 43 ATTAAGCTTCGGCGGAGCGCGCTCGCCCTCCACTCCGCCAGCTCCGGGAGAGGAG 102  
Qy 61 CCGCACCG 120  
Db 103 CCGCACCG 162  
Qy 121 TGTAGCGTGCCTGTGCACGGTGTCTGGCACCTCCCTTCAGTGCAGCTCTCAATCGCA 180  
Db 163 TGTAGCGTGCCTGTGCACGGTGTCTGGCACCTCCCTTCAGTGCAGCTCTCAATCGCA 222  
Qy 181 CAGGGGACCCAGGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 223 CAGGGGACCCAGGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 282  
Qy 241 AGCAACCACTCCGACGCTTGGCTGGGACTATCTCAAACTACCTGGGCGCGCGCGCG 300  
Db 283 AGCAACCACTCCGACGCTTGGCTGGGACTATCTCAAACTACCTGGGCGCGCGCGCG 342  
Qy 301 AGCAACCACTCCGACGCTTGGCTGGGCGCGAGAGACTTGCCCGAGGGCGACTGTGACT 360  
Db 343 AGCAACCACTCCGACGCTTGGCTGGGCGCGAGAGACTTGCCCGAGGGCGACTGTGACT 402  
Qy 361 TGGAGGTGGGGAAGCCCTCAATGACAAACTGGGGTGAAGCCAGAACTAGAGGCTTACA 420  
Db 403 TGGAGGTGGGGAAGCCCTCAATGACAAACTGGGGTGAAGCCAGAACTAGAGGCTTACA 462  
Qy 421 GCCACCTTCTGTGTTACTTGGCTGCGCTCAACCGTCAGGCTGCCACTGTGAGCTGGGCC 480  
Db 463 GCCACCTTCTGTGTTACTTGGCTGCGCTCAACCGTCAGGCTGCCACTGTGAGCTGGGCC 522  
Qy 481 GCAGCTGGCCCACTTCTGACACAGCTCCAGGGCTGCTGGGAGCAATTCGGGGCTTCA 540  
Db 523 GCAGCTGGCCCACTTCTGACACAGCTCCAGGGCTGCTGGGAGCAATTCGGGGCTTCA 582  
Qy 541 TGGCAGCTTGGGCTACCCACTGCCCGAGCGCTGCCCTGGGAGCTGAACCCACTTGGACTC 600  
Db 583 TGGCAGCTTGGGCTACCCACTGCCCGAGCGCTGCCCTGGGAGCTGAACCCACTTGGACTC 642

QY 601 CTGGCCCTGCCACAGTGAATCTCTCCAGAGATGACGACTTCTGGCTGCTGAAGAGC 660  
 Db 643 CTGGCCCTGCCACAGTGAATCTCTCCAGAGATGACGACTTCTGGCTGCTGAAGAGC 702  
 QY 661 TGCAGACCTGGCTGTGGCGCTGGCCCAAGGACTTCAACCGGCTCAAGAGAGATGACG 720  
 Db 703 TGCAGACCTGGCTGTGGCGCTGGCCCAAGGACTTCAACCGGCTCAAGAGAGATGACG 762  
 QY 721 CTCAGACCTGAGTACCTGCACTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780  
 Db 763 CTCAGACCTGAGTACCTGCACTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 822  
 QY 781 CCTCTTGGCTCCCC 795  
 Db 823 CCTCTTGGCTCCCC 837

## RESULT 7

AAH99772  
 ID AAH99772 standard; cDNA; 768 BP.  
 AC AAH99772;  
 AT 16-OCT-2001 (first entry)

Human protein encoding cDNA sequence SEQ ID NO:607.

Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 antibacterial; endocrine; cardiant; central nervous system; virucide;  
 anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;  
 dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
 neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 antinaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 allergic rhinitis; diabetes; multiple sclerosis; depression;  
 Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 neurological disorder; ss.

Homo sapiens.

WO200153455-A2.

26-JUL-2001.

22-DEC-2000; 2000WO-US35017.

23-DEC-1999; 99US-0471275.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-457603/49.

P-PSDB; AAM25831.

Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers, and HIV infection -

Claim 1; Page 638; 1217pp; English.

AAH99166 to AAH99904 encode the human proteins given in AAM25225 to AAM25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antiinflammatory; antirheumatic; antiallergic; immunosuppressive; antibacterial; endocrine; cardiant; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;

CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,  
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders.

Sequence 768 BP; 137 A; 278 C; 217 G; 136 T; 0 other;

Query Match 95.3%; Score 759.2; DB 22; Length 768;  
 Best Local Similarity 99.6%; Pred. No. 7.3e-168;  
 Matches 761; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATTAAGCTTCGCGAGCGCGCTCGCCCTCCCACTCCGCGAGCTCCGCGAGGAG 60  
 Db 5 ATTAAGCTTCGCGAGCGCGCTCGCTCTTCACTCCGCGAGCTCCGCGAGGAG 64  
 QY 61 CCGCACCG 120  
 Db 65 CCGCACCG 124  
 QY 121 TGTAGGCTGTCTGTGACAGCGTGTCTTGGCACCTCCCTGCGAGTCCGAGCTCTCAATCGCA 180  
 Db 125 TGTAGGCTGTCTGTGACAGCGTGTCTTGGCACCTCCCTGCGAGTCCGAGCTCTCAATCGCA 184  
 QY 181 CAGGGACCGAGCGCTGGCGCTTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240  
 Db 185 CAGGGACCGAGCGCTGGCGCTTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 244  
 QY 241 AGCACCAACTCCGAGCTTGGCTGGGACCTTCTGAACCTACCTGGGCGCGCTTTCAAGC 300  
 Db 245 AGCACCAACTCCGAGCTTGGCTGGGACCTTCTGAACCTACCTGGGCGCGCTTTCAAGC 304  
 QY 301 AGCCAGCTTCAACCTCCCGCGCTGGGGCGAGAGCTCTGCCAGGGCGCACTGTTGACT 360  
 Db 305 AGCCAGCTTCAACCTCCCGCGCTGGGGCGAGAGCTCTGCCAGGGCGCACTGTTGACT 364  
 QY 361 TGGAGCTGTGGCGAAGCGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGCGCTTACA 420  
 Db 365 TGGAGCTGTGGCGAAGCGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGCGCTTACA 424  
 QY 421 GCCACCTTCTGTGTTACTTGTGGCTCAACCGTCAGGCTGCACCTTGGGGCGCTTCAAGC 480  
 Db 425 GCCACCTTCTGTGTTACTTGTGGCTCAACCGTCAGGCTGCACCTTGGGGCGCTTCAAGC 484  
 QY 481 GCAGCTTGGCGCACTTCTGCAACAGCTCCAGGGCTTCTGGGGCGCACTTGGGGCGCTTCA 540  
 Db 485 GCAGCTTGGCGCACTTCTGCAACAGCTCCAGGGCTTCTGGGGCGCACTTGGGGCGCTTCA 544  
 QY 541 TGGCAGCTCTGGGCTACCCACTGCCCCCGCGCGCTGCTGGGCGAGCTTGAACCTACCTGACTC 600  
 Db 545 TGGCAGCTCTGGGCTACCCACTGCCCCCGCGCGCTGCTGGGCGAGCTTGAACCTACCTGACTC 604  
 QY 601 CTGGCGCTGCCACAGTGAATCTCTCCAGAGATGACGACTTCTGGCTGCTGAAGAGC 660  
 Db 605 CTGGCGCTGCCACAGTGAATCTCTCCAGAGATGACGACTTCTGGCTGCTGAAGAGC 664  
 QY 661 TGCAGACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGACG 720  
 Db 665 TGCAGACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGACG 724  
 QY 721 CTCAGACCTGAGTACCTGCACTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 764  
 Db 725 CTCAGACCTGAGTACCTGCACTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 768



FT /product= "Human mature cardiostrophin-like cytokine  
 XX (CLC) protein"  
 PN WO200127157-A1.  
 XX 19-APR-2001.  
 XX 06-OCT-2000; 2000WO-AU01216.  
 XX 08-OCT-1999; 99AU-0003327.  
 PR 12-MAY-2000; 2000AU-0007489.  
 XX (AMRA-) AMRAD OPERATIONS PTY LTD.  
 XX Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;  
 PI Nakata Y, Hasegawa M;  
 DR P-PSDB; AAE00828.  
 XX  
 XX New biologically active complex comprising NR6 and  
 PT cardiostrophin-like-cytokine, for facilitating proliferation,  
 PT differentiation and/or survival of a cell -  
 XX  
 PS Claim 31; Page 112-114; 123pp; English.  
 XX  
 CC The present invention relates to a biologically active complex comprising  
 CC a haemopoietin receptor, NR6 and cardiostrophin-like cytokine (CLC).  
 CC The complex is useful in the manufacture of a medicament for the  
 CC treatment and/or prophylaxis of a subject, as it is involved in  
 CC facilitating proliferation, differentiation and/or survival of a cell.  
 CC The complex or its components have neurotrophic activity. The present  
 CC sequence is human cardiostrophin-like cytokine (CLC) cDNA.  
 XX  
 SQ Sequence 729 BP; 132 A; 261 C; 196 G; 140 T; 0 other;  
 Query Match 89.1%; Score 710; DB 22; Length 729;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-156;  
 Matches 710; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 86 CCCATGACCTCCGAGCAGGGAGCTGCTGGGGGATGTTAGCGTGCCTGTCACGGTGTCT 145  
 DB 7 CCCATGAGCCTCCGAGCAGGGAGCTGCTGGGGGATGTTAGCGTGCCTGTCACGGTGTCT 66  
 QY 146 CTGGACCTCCCTGACGTCAGTCCAGCTCTCAATCGCAGGGGACCCAGGGCTGGCCCCCTC 205  
 DB 67 CTGGACCTCCCTGACGTCAGTCCAGCTCTCAATCGCAGGGGACCCAGGGCTGGCCCCCTC 126  
 QY 206 CATCAGAAAACTATGACCTCACCGCTACCTGGAGCACCACCACTCCGAGCTTGGCTGG 265  
 DB 127 CATCAGAAAACTATGACCTCACCGCTACCTGGAGCACCACCACTCCGAGCTTGGCTGG 186  
 QY 266 GACCTATCTGAACCTGCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCT 325  
 DB 187 GACCTATCTGAACCTGCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCT 246  
 QY 326 GGGGCGAGAGCTCTGCCAGGGCCACTGTTGACTTGGAGGTGGCGAAGCTCAATGA 385  
 DB 247 GGGGCGAGAGCTCTGCCAGGGCCACTGTTGACTTGGAGGTGGCGAAGCTCAATGA 306  
 QY 386 CAAACTGGGCTGACCCAGAACTAGAGGCTTACAGCCACTTCTGTGTTACTTGGCTGG 445  
 DB 307 CAAACTGGGCTGACCCAGAACTAGAGGCTTACAGCCACTTCTGTGTTACTTGGCTGG 366  
 QY 446 CCTCAACCGTTCAGGCTGACCTGCTGAGCTGGCGGCGAGCTGGGCCCTCTTGTGACCCAG 505  
 DB 367 CCTCAACCGTTCAGGCTGACCTGCTGAGCTGGCGGCGAGCTGGGCCCTCTTGTGACCCAG 426  
 QY 506 CCTCAGGGCTGCTGGGAGCATTGGCGGCGTCAATGGCAGCTCTGGGCTACCCACTGCC 565  
 DB 427 CCTCAGGGCTGCTGGGAGCATTGGCGGCGTCAATGGCAGCTCTGGGCTACCCACTGCC 486  
 QY 566 CCAGCGCTGCTGGGAGCTGAACCACTTGGACTCTCTGGGCCCTCCCACTGACTTCTCT 625

Db 487 CCAGCGCTGCTGGAGTGAACCCACTTGGACTCTCTGGCCCTGCCACAGTGAATCTCT 546  
 QY 626 CCAGAAGATGACGACTTCTGGCTGCTGAAGAGCTGAGACCTGGCTGGCTGGCTGGCTGG 685  
 Db 547 CCAGAAGATGACGACTTCTGGCTGCTGAAGAGCTGAGACCTGGCTGGCTGGCTGGCTGG 606  
 QY 686 CAAGGACTTCAACCGGCTCAAGAGAGATGACGCTCCAGCAGCTGACCTGCACTGCA 745  
 Db 607 CAAGGACTTCAACCGGCTCAAGAGAGATGACGCTCCAGCAGCTGACCTGCACTGCA 666  
 QY 746 CCTGGGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCT 795  
 Db 667 CCTGGGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCT 716  
 RESULT 10  
 AAK51548  
 ID AAK51548 standard; cDNA; 1008 BP.  
 XX  
 AC AAK51548;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 93.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US04098.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR P-PSDB; AAM78415.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 PS Claim 1; Page 711-712; 622pp; English.  
 CC  
 CC The invention relates to polynucleotides (AAK51456-AAK51435) and the  
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.



CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAW80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

[illegible]

RESULT 11  
AAV47512  
ID AAV47512 standard: CDNA: 819 BP.

XX 09-NOV-1998 (first entry)  
DT  
XX Manager: Neurotrophin factor NNT-1 CDNA

NNT-1; neurotrophic factor; mouse; antinflammatory; adjuvant; Alzheimer's disease; Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; peripheral neuropathy; dystrophy; neural retina degeneration; common variable immunodeficiency; CVID; selective IgA deficiency; hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic; the x-ray. 55

	Mus SP.	Location/Qualifiers
Key	95..772	/tag= a
CDS	95..175	/tag= b
sig_peptide	176..769	/tag= c
mat_peptide		
FT		
XX	WO9833922-A1.	
XX		
XX	06-AUG-1998.	
XX		
XX	02-FEB-1998;	98WO-US02363.
XX		
XX	30-JAN-1998;	98US-0016534.
XX	03-FEB-1997;	97US-0792019.
XX	(AMGE-) AMGEN INC.	
XX		
XX	Chang M, Elliot GS, Sarmiento U,	
XX		
XX	WPI; 1998-437475/37.	
XX	P-PSDB: AAW29716.	
XX		

XX Newly isolated nucleic acid encoding human or murine neurotrophic  
XX factor NNT-1 - useful for treatment of neurological and  
XX immunological diseases or inflammation, also as vaccine adjuvant  
XX Disclasure: fig 4. 120pp: English.

This newly isolated mouse cDNA sequence codes for a novel neurotrophic factor, designated NNT-1 (see AW29716), that is a growth factor for neurons and for B or T cells. Vectors and host cells are provided for use in the production of murine and human NNT-1 polypeptides. These are used to treat: (i) neurological or immunological diseases, specifically Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and degeneration of the neural retina, or conditions characterised by T or B cell defects, e.g. common variable immunodeficiency (CVID), selective IgA deficiency, hypogammaglobulinaemia and X-linked agammaglobulinaemia (claimed), but many others disclosed; and (ii) inflammation. NNT-1 is also able to boost immunoreactivity and antibody production following vaccination, and, since it inhibits tumour necrosis factor production, it may also be useful for treating sepsis. NNT-1 nucleic acid fragments are also used as hybridisation probes in diagnostic assays. In addition, cells that have been engineered to express NNT-1 can be implanted, or nucleic acids are delivered in gene therapy vectors.

Sequence 819 BP: 156 A; 288 C; 218 G; 157 T; 0 other;

Query Match	84.0%;	Score 669.4;	DB 19;	Length 819;
Best Local Similarity	92.0%;	Pred. No. 6.8e-147;		
		Mismatches 61;	Indels 1;	Gaps 1;

Qy	1	ATTAAAGCTTCGCGGAGCGCGCTCGCTCCCTCCCACTCGCCAGCCTCCGGGAGAGGAG	60
Db	5	ATTAAAGCTTCGCGGAGCGCGCTCGCTCCCTCCCACTCGCCAGCCTTCGGGAGAGGAG	64
Qy	61	CCGACACCGGCGCGGCCGAG-CCCAAGCTCCATGCACTTCGAGCAGGGGACTCGTGGGGC	11
Db	65	CCGCGCGGCGCGCGCGGCCCGCCAGCCCCATGACCTTCGAGCAGGGGACTCGTGGGGC	12



```

QY 120 ATGTTAGCGTCTGTGCGAGCGTGTCTGCGACCTCCCTGCGAGTGCAGCTCTCAATCGC 179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
125 ATGTTAGCGTCTGTGCGAGCGTGTCTGCGACCTCCCTGCGAGTGCAGCTCTCAATCGC 184
QY 180 ACAGGGGACCCAGGGGCTGGCCCTCCATCCAGAAAACCTATGACCTACCCGCTACCTG 239
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
185 ACAGGAGATCCAGGGCTGGCCCTCCATCCAGAAAACCTATGACCTACCCGCTACCTG 244
QY 240 GAGCACCACCTCGCAGCTGTGGAGCTATCTGAACTACTCTGGGCCCCCTTTCAAC 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
245 GAGCATCACTCCGAGCTTAGCTGGAGCTACCTGAACTACTCTGGGCCCCCTTTCAAC 304
QY 300 GAGCCAGACTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTGTGAC 359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
305 GAGCTGACTTCAATCTCTCGACTGGGGGAGAAAACCTCTGCCAGGGCCACGCTCAAC 364
QY 360 TTGGAGGTGTGGCGAAGCTCAATGACAACTCGCGGCTGACCCAGAACTACGAGCGCTAC 419
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365 TTGGAAGTGTGGCGAAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGCGTAC 424
QY 420 AGCCACCTCTGTGTACTTGGTGGCTCAACCGTCAAGGCTGCCACTGTGAGCTGGC 479
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
425 AGTCACCTCTGTGTACTTGGTGGCTCAACCGTCAAGGCTGCCACTGTGAGCTGGC 484
QY 480 CGCAGCTTGGCCCACTCTGACACAGCTCCAGGGCTGCTGGGAGCAATTCGAGGTGC 539
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485 CGTAGCTTGGCCCACTCTGACACAGCTCCAGGGCTGCTGGGAGCAATTCGAGGTGC 544
QY 540 ATGCGAGCTTGGGCTACCCACTGCCAGCGCTGCTGGGAGCAATTCGAGGTGC 599
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545 ATGCGAGCTTGGGCTACCCACTGCCAGCGCTGCTGGGAGCAATTCGAGGTGC 604
QY 600 CTTGGCCCTTGCACAGTACTTCTCCAGAGATGAGAGCTTCTGGCTGTGAAGGAG 659
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
605 CTTGGCCCTTGCACAGTACTTCTCCAGAGATGAGAGCTTCTGGCTGTGAAGGAG 664
QY 660 CTGCGAGCTTGGCTGTGGCGCTGGCCAAAGACTTCAACCGGCTCAAGAGAGATGCGAG 719
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
665 CTGCGAGCTTGGCTATGGGCTTCAAGCAAGGACTTCAACCGGCTTAAAGAAAGATGCGAG 724
QY 720 CTTCCAGAGCTGAGTACCTTGCAGCTGGGGCTCATGGCTTCTGACCTTCTGACCTT 778
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
725 CTTCCAGAGCTTCACTGCTGACCTTGGAGGCAATGGTTTCTGACCTTCTGACCTT 783

RESULT 12
AAV22654
ID AAV22654 standard; cDNA; 819 BP.
AC AAV22654;
DT 13-JUL-1998 (first entry)
DE cDNA encoding murine neurotrophic factor NNT-1.
KW Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
KW treatment; neurological disease; degeneration; Parkinson's disease;
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke; ss.
OS Mus sp.
FH Key
FT CDS 95..771 Location/Qualifiers
FT sig_peptide /tag= a
FT mat_peptide /tag= b
FT /tag= c
US5741772-A.
21-APR-1998.

```

```

XX 03-FEB-1997; 97US-0792019.
XX 03-FEB-1997; 97US-0792019.
XX (AMGE-) AMGEN INC.
XX Chang M;
XX WPI; 1998-260526/23.
XX P-PSDB; AAW56142.
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
XX useful for stimulating growth of motor and sympathetic neurons
XX Disclosure; Fig 4; 41pp; English.
XX The present sequence encodes a murine neurotrophic factor, designated
XX NNT-1, which is capable of stimulating growth of motor or sympathetic
XX neurons. The NNT-1 protein is useful in the treatment of neurological
XX diseases characterized by the degeneration and death of particular
XX classes of neurons. These diseases specifically include Parkinson's
XX disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
XX stroke and various degenerative disorders affecting vision.
XX Sequence 819 BP; 156 A; 288 C; 218 G; 157 T; 0 other;
Query Match 84.0%; Score 669.4; DB 19; Length 819;
Best Local Similarity 92.0%; Pred. No. 6.8e-147;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
QY 1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACTCCGCCAGCTTCGGGAGAGAG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACTCCGCCAGCTTCGGGAGAGAG 64
QY 61 CCGCACCCGCGCGCGCCAG-CCCGAGCCCATGAGACCTCCGAGCAGGGAGCTCGTGGGG 119
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
65 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 124
QY 120 ATCTTACGCTGCTGTGCGAGCTGTCTGGCACTCCCTGCGAGTGCAGCTCTCAATCGC 179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
125 ATCTTACGCTGCTGTGCGAGCTGTCTGGCACTCCCTGCGAGTGCAGCTCTCAATCGC 184
QY 180 ACAGGGGACCCAGGGCTTGGCTGGGAGCTATCTGAACTACCTGGGCCCCCTTTCAAC 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
185 ACAGGAGATCCAGGGCTTGGCCCTCCATCCAGAAAACCTATGACCTACCCGCTACCTG 244
QY 240 GAGCACCACCTCGCAGCTTGGCTGGGAGCTATCTGAACTACCTGGGCCCCCTTTCAAC 304
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
245 GAGCATCACTCCGAGCTTAGCTGGAGCTACCTGAACTACTCTGGGCCCCCTTTCAAC 304
QY 300 GAGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTGTGAC 359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
305 GAGCTGACTTCAATCTCTCGACTGGGGGAGAAAACCTCTGCCAGGGCCACGCTCAAC 364
QY 360 TTGGAGGTGTGGCGAAGCTCAATGACAACTCGCGGCTGACCCAGAACTACGAGCGCTAC 419
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365 TTGGAAGTGTGGCGAAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGCGTAC 424
QY 420 AGCCACCTCTGTGTACTTGGTGGCTCAACCGTCAAGGCTGCCACTGTGAGCTGGC 479
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
425 AGTCACCTCTGTGTACTTGGTGGCTCAACCGTCAAGGCTGCCACTGTGAGCTGGC 484
QY 480 CGCAGCTTGGCCCACTCTGACACAGCTCCAGGGCTGCTGGGAGCAATTCGAGGTGC 539
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485 CGTAGCTTGGCCCACTCTGACACAGCTCCAGGGCTGCTGGGAGCAATTCGAGGTGC 544
QY 540 ATGCGAGCTTGGGCTACCCACTGCCAGCGCTGCTGGGAGCAATTCGAGGTGC 599
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545 ATGCGAGCTTGGGCTACCCACTGCCAGCGCTGCTGGGAGCAATTCGAGGTGC 604
QY 600 CTTGGCCCTTGCACAGTACTTCTCCAGAGATGAGAGCTTCTGGCTGTGAAGGAG 659

```



KW Mouse; ss; gene; NNT1; neurotrophic factor; IgE-related disease;  
 KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;  
 KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;  
 KW vascular restenosis; rheumatoid arthritis; psoriatic arthritis;  
 KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;  
 KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;  
 KW inflammatory bowel disease; transplant rejection; reproductive disorder;  
 KW graft versus host disease; infertility; miscarriage; preterm labour.  
 OS Mus sp.

Key Location/Qualifiers  
 CDS 95..772

FT /\*tag= a  
 FT /product= "NNT1"

PN WO200215977-A2.

XX 28-FEB-2002.

XX 17-AUG-2001; 2001WO-US25906.

XX 18-AUG-2000; 2000US-226436P.

XX 16-AUG-2001; 2001US-0931704.

XX (AMGE-) AMGEN INC.

PA Senaldi G;

XX WPI; 2002-280867/32.

XX P-PSDB; AAU78177.

XX Treating Immunoglobulin E-related disease, modulating IgE levels in a  
 XX patient, preventing IgE-related disease and treating allergic diseases,  
 XX involves administering NNT-1 inhibitor to a patient  
 XX Claim 2; Fig 4; 63pp; English.

XX The invention relates to treating Immunoglobulin E (IgE)-related disease,  
 XX modulating IgE levels in a patient, preventing an IgE-related disease,  
 XX and treating allergic diseases, comprising administering a  
 XX therapeutically effective amount of novel neurotrophic factor (NNT)-1  
 XX inhibitor to a patient. Also included are a method of diagnosing an  
 XX IgE-related disease or susceptibility to an IgE-related disease, by  
 XX determining the presence or amount of expression of an NNT1 polypeptide  
 XX encoded by a NNT1 nucleotide sequence, its fragment or naturally  
 XX occurring variant, and diagnosing an IgE-related disease or  
 XX susceptibility of an IgE-related disease based on the presence or amount  
 XX of expression of the polypeptide and a pharmaceutical composition for use  
 XX in treating IgE-related disease, comprising the NNT1 inhibitor.  
 XX The NNT1 inhibitor is useful for preventing and treating IgE-related  
 XX disease, modulating IgE levels, and treating allergic diseases e.g.  
 XX Type I allergic disease, allergic rhinitis, eczema, dermatitis,  
 XX pollinosis, asthma, immune diseases and disorders, diseases involving  
 XX abnormal cell proliferation including cancer, arteriosclerosis and  
 XX immune system including rheumatoid arthritis, psoriatic arthritis,  
 XX inflammatory arthritis, osteoarthritis, inflammatory joint disease,  
 XX autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory  
 XX bowel disease, transplant rejection, and graft versus host disease, and  
 XX reproductive diseases and disorders including infertility, miscarriage,  
 XX preterm labour and delivery, and endometriosis. The present sequence  
 XX encodes mouse NNT1.

XX Sequence 819 BP; 156 A; 288 C; 218 G; 157 T; 0 other;

Query Match 84.0%; Score 669.4; DB 24; Length 819;  
 Best Local Similarity 92.0%; Pred. No. 6.8e-147;  
 Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

QY 1 ATTAAGCTTCGCGGAGCCGGCTCGCCCTCCACCTCCGCGAGCCCTCCGGAGAGGAG 60  
 DB 5 ATTAAGCTTCGCGGAGCCGGCTCGCCCTCCACCTCCGCGAGCCCTTCGGAGAGGAG 64

QY 61 CGCAGCCCGCGCGCCAG-CCCCAGCCCATGAGCTTCGAGCAGGGGACTCGTGGGG 119  
 DB 65 CG 124  
 QY 120 ATGTTAGCGTGTGTCAGCGGTGCTCTGGACCTCTCCCTGTCAGTGCAGCTCTCAATCGC 179  
 DB 125 ATGTTAGCTTGCTATGACACGGTGTGTGGACCTCTCCCTGTCAGTGCAGCTCTTAATCGC 184  
 QY 180 ACAGGGAGCCAGGGCGCTGGCGCCCTCCATCCAGAAAACCTATGACCTACCGCTACCTG 239  
 DB 185 ACAGGAGATCCAGGCGCTGGCGCCCTCCATCCAGAAAACCTATGACCTACCGCTACCTG 244  
 QY 240 GAGCAGCACTCCGCGAGCTTGCTGGACCTATCTGAATACCTGGCGCCCGCTTTCAAC 299  
 DB 245 GAGCATCAACTCCGCGAGCTTAGCTGGACCTACCTGAACTACCTGGCGCCCGCTTTCAAC 304  
 QY 300 GAGCCAGACTTCAACCCCTCCCGCTGGCGGCGAGAGCTCTGCCAGGGGCACTGTTGAC 359  
 DB 305 GAGCTGACTTCAATCTCTCGACTGGGGGAGAAAATCTGCGCCAGGGGCCACGGTCAAC 364  
 QY 360 TTGAGGTGTGGGAGGCTCAATGACAAACTGCGGCTGACCCAGAACTACGAGGCTTAC 419  
 DB 365 TTGGAAGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGCGTAC 424  
 QY 420 AGCCACCTTCTGTCTTACTTGGCTGGCTCAACCGTACGGCTGCACCTGCTGAGTGGC 479  
 DB 425 AGTCACCTCTGTGTGTACTTGGCTGGCTCAACCGTACGGCTGCACCTGCTGAGTGGC 484  
 QY 480 CGCAGCTGGCCCACTTCTGACCCAGGCTCCAGGGCGCTGCTGGGCGAGCATTCGGGCGTC 539  
 DB 485 CGTAGCTGGCCCACTTCTGACCGGCTCCAGGGCGCTGCTGGGCGAGCATTCGGGCGTC 544  
 QY 540 ATGCGAGCTTGGGCTACCCACCTGCCAGCGCTGCTGGGAGTGAACCCACCTTGGACT 599  
 DB 545 ATGGCGAGCTTGGCTACCCACCTGCCAGCGCTGCTGGGAGTGAACCCACCTTGGACT 604  
 QY 600 CTGGCGCTGCCCACTGAGTCTTCTCCAGAGATGACGACTTCTGGCTGCTGAAGAG 659  
 DB 605 CTGGCGCTGCCCACTGAGTCTTCTCCAGAGATGAGTACTTCTGGCTGCTGAAGAG 664  
 QY 660 CTCGAGCTGGCTGGCGCTCGGCGCTCGGCGAGTCTCAACCGGCTCAAGAGAGATGAG 719  
 DB 665 CTCGAGCTGGCTGGCGCTCGGCGCTCGGCGAGTCTCAACCGGCTTAAGAGAGATGAG 724  
 QY 720 CTCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 778  
 DB 725 CTCCAGAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 783

RESULT 15

ABA09140

ID ABA09140 standard; cDNA; 968 BP.

XX ABA09140;

XX 11-JAN-2002 (first entry)

DE Human cardiostrophin-like cytokine homologue cDNA, SEQ ID NO:916.

KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; proliferative disorder;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;  
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cyostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnary; antiulcer; 98.

us-09-931-704-1.rng

Sun Feb 2 08:31:37 2003

XX OS Homo sapiens.  
XX PN WO200157188-A2.  
XX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US03800.  
XX PR 03-FEB-2000; 2000US-0496914.  
XX PR 27-APR-2000; 2000US-0560875.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Liu C, Drmanac RT;  
XX PD WPI; 2001-457740/49.  
XX DR P-PSDB; ABB11896.  
XX PT Human proteins and DNA encoding sequences useful for preventing,  
XX PT treating or ameliorating a medical condition in a mammalian subject  
XX PT e.g. arthritis and cancer -  
XX PS Claim 1; Page 793-794; 1963pp; English.  
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
XX invention also relates to vectors and recombinant host cells comprising a  
XX nucleotide of the invention, methods of producing the novel polypeptides,  
XX antibodies against the polypeptides, methods of detecting the nucleotides  
XX or polypeptides in a sample, and methods of identifying compounds which  
XX bind to polypeptides of the invention. Although novel, many of the  
XX polypeptides of the invention have homology to known proteins, and thereby  
XX giving an insight into their probable biological activities, and hence  
XX potential therapeutic applications. The polypeptides of the invention may  
XX have various activities, including cytokine, cell proliferation or cell  
XX differentiation activities; stem cell growth factor activity;  
XX haematopoiesis regulatory activity; tissue growth activity;  
XX immunomodulatory activity; activin- or inhibin-related activities;  
XX chemotactic or chemokinetic activities; haemostatic, thrombotic or  
XX thrombolytic activities; receptor or ligand activities; or may be  
XX involved in oncogenesis, cancer cell proliferation or metastasis.  
XX Depending on their biological activities, polypeptides and nucleotides of  
XX the invention are useful for preventing, treating or ameliorating medical  
XX conditions, e.g., by protein or gene therapy. Such conditions include  
XX cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
XX disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
XX proliferative retinopathy, atherosclerosis, coronary heart disease,  
XX arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
XX vascular growth. Polypeptides involved with tissue regeneration and  
XX repair (or nucleic acids encoding them) may be used to promote wound  
XX healing (e.g., of burns, incisions and ulcers), while those with  
XX immunomodulatory activities may be used in the treatment of viral,  
XX bacterial and fungal infections in addition to immune disorders.  
XX Polypeptides with growth factor activity may be used in cell cultures to  
XX promote cell growth. For example, such polypeptides may be used to  
XX manipulate stem cells in culture to give rise to neuroepithelial cells  
XX that can be used to augment or replace cells damaged by illness,  
XX autoimmune disease or accidental damage. The polypeptides and nucleotides  
XX may also be used in the diagnosis of the above conditions, and in drug  
XX screening techniques. The present sequence represents a cDNA encoding a  
XX novel human polypeptide of the invention.

XX SQ Sequence 968 BP; 179 A; 353 C; 244 G; 192 T; 0 other;

Query Match 83.9%; Score 668.4; DB 22; Length 968;  
Best Local Similarity 97.0%; Pred. No. 1.2e-146;  
Matches 681; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

XX 66 'CCCCGGCCGCGCCAGCCCGCCATGAGCTCCGAGCAGGGGACTCGTGGGGGATGTTA 125  
DB 267 CCGGGCCGCGCTCATGCTCTCTGCTCCATCTGCCACAGGGGAGCTGCTGGGGGATGTTA 326

QY 126 GCGTGCCTGTGTCACGGTGCTCTGGCACCTCCCTGTCAGTGCAGCTCTCAATGCACAGGG 185  
DB 327 GCGTGCCTGTGTCACGGTGCTCTGGCACCTCCCTGTCAGTGCAGCTCTCAATGCACAGGG 386  
QY 186 GACCCAGGGCCTGGCCCTCCATCCAGAAAACCTATGACCTCACCGCTACCTGGAGCAC 245  
DB 387 GACCCAGGGCCTGGCCCTCCATCCAGAAAACCTATGACCTCACCGCTACCTGGAGCAC 446  
QY 246 CAATCCCGCAGCTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCA 305  
DB 447 CAATCCCGCAGCTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCA 506  
QY 306 GACTTCAACCTCCCGCTGGGCGCAGAGACTCTGCCAGGGCCACTGTTGACTTTGGAG 365  
DB 507 GACTTCAACCTCCCGCTGGGCGCAGAGACTCTGCCAGGGCCACTGTTGACTTTGGAG 566  
QY 366 GTGTGGCGAAGCCTCAATGACAAAACCTGCGGCTGACCCAGAACTACGAGGCTTACAGCCAC 425  
DB 567 GTGTGGCGAAGCCTCAATGACAAAACCTGCGGCTGACCCAGAACTACGAGGCTTACAGCCAC 626  
QY 426 CTTCTGTGTACTTGTGCTGGCCCTCAACCGTCAAGCTGAGCTGCCACTGCTGAGCTGGCGCGCAGC 485  
DB 627 CTTCTGTGTACTTGTGCTGGCCCTCAACCGTCAAGCTGAGCTGCCACTGCTGAGCTGGCGCGCAGC 686  
QY 486 CTGGCCCACTTCTGCAACCAAGCTCCAGGGCTGCTGGGCGCATTTGGGGGCTCATGGCA 545  
DB 687 CTGGCCCACTTCTGCAACCAAGCTCCAGGGCTGCTGGGCGCATTTGGGGGCTCATGGCA 746  
QY 546 GCTCTGGGCTTACCCACTGCCCCAGCGCTGCTGGGACTGAACCCACTTTGGACTCTCTGGC 605  
DB 747 GCTCTGGGCTTACCCACTGCCCCAGCGCTGCTGGGACTGAACCCACTTTGGACTCTCTGGC 806  
QY 606 CTTGGCCACAGTGAATTCCTCCAGAACTGACCACTTCTGGCTGCTCAAGGAGCTGCAG 665  
DB 807 CTTGGCCACAGTGAATTCCTCCAGAACTGACCACTTCTGGCTGCTCAAGGAGCTGCAG 866  
QY 666 ACCTGGCTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAAAGAGATGAGCTTCCA 725  
DB 867 ACCTGGCTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAAAGAGATGAGCTTCCA 926  
QY 726 GCAGCTGCAGTCAACCTGCACTGGGGGCTCATGGCTTCTGA 767  
DB 927 GCAGCTGCAGTCAACCTGCACTGGGGGCTCATGGCTTCTGA 968

Search completed: February 1, 2003, 09:11:18  
Job time : 206.966 secs



```
Db 121 TGTTAGCGTCTGTGACAGGTGCTCTGGCACCTCCCTGCGAGTCCGACGCTCTCAATCGCA 180
Qy 181 CAGGGAGCCAGGCGCTGGCCCTCCATCCAGAAAACTATGACTCAACCGCTACCTGG 240
Db 181 CAGGGAGCCAGGCGCTGGCCCTCCATCCAGAAAACTATGACTCAACCGCTACCTGG 240
Qy 241 AGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 300
Db 241 AGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 300
Qy 301 AGCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGCGCACTGTTGACT 360
Db 301 AGCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGCGCACTGTTGACT 360
Qy 361 TGAGGTGTGCGAAGCTCAATGACAACTGCGGCTGACCAAGCTACAGAGCTTACA 420
Db 361 TGAGGTGTGCGAAGCTCAATGACAACTGCGGCTGACCAAGCTACAGAGCTTACA 420
Qy 421 GCCACCTTCTGTGTTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCGCC 480
Db 421 GCCACCTTCTGTGTTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCGCC 480
Qy 481 GCAGCTTGGCCACTTGTGACAGGCTTCCAGGCGCTGCTGGGAGGANTGGGGCGTCA 540
Db 481 GCAGCTTGGCCACTTGTGACAGGCTTCCAGGCGCTGCTGGGAGGANTGGGGCGTCA 540
Qy 541 TGGCAGCTTGGGCTACCCACTGCCCGCGCTGCCCTGGGACTGAACCCACTTGGACTC 600
Db 541 TGGCAGCTTGGGCTACCCACTGCCCGCGCTGCCCTGGGACTGAACCCACTTGGACTC 600
Qy 601 CTGGCCCTGCCACAGTGACTTCTCCAGAAAGATGGACGACTTCTGGCTGTGAAGGAGC 660
Db 601 CTGGCCCTGCCACAGTGACTTCTCCAGAAAGATGGACGACTTCTGGCTGTGAAGGAGC 660
Qy 661 TGACAGCTTGGCTGTGGCGCTCGGCCAAGAGCTTCAACCGGCTCAAGAAAGATGAGC 720
Db 661 TGACAGCTTGGCTGTGGCGCTCGGCCAAGAGCTTCAACCGGCTCAAGAAAGATGAGC 720
Qy 721 CTCAGAGCTGACGTCAACCTGCACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
Db 721 CTCAGAGCTGACGTCAACCTGCACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
Qy 781 CTTCTTGGCTCCCCC 797
Db 781 CTTCTTGGCTCCCCC 797
```

## RESULT 2

US-09-931-704-4

; Sequence 4, Application US/09931704

; Patent No. US20020041873A1

; GENERAL INFORMATION:

; APPLICANT: Senaldi, Giorgio

; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using N

; FILE REFERENCE: A-695

; CURRENT APPLICATION NUMBER: US/09/931,704

; PRIOR FILING DATE: 2001-08-16

; PRIOR APPLICATION NUMBER: US 60/226,436

; PRIOR FILING DATE: 2000-08-18

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 819

; TYPE: DNA

; ORGANISM: Murine

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (95)..(769)

; OTHER INFORMATION:

; NAME/KEY: mat\_peptide

; LOCATION: (176)..()

; OTHER INFORMATION:

```
; NAME/KEY: mat_peptide
; LOCATION: (176)..(769)
; OTHER INFORMATION:
; NAME/KEY: sig_peptide
; LOCATION: (95)..(175)
; OTHER INFORMATION:
US-09-931-704-4
```

Query Match 84.0%; Score 669.4; DB 10; Length 819;

Best Local Similarity 92.0%; Pred. No. 1.1e-150;

Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

```
Qy 1 ATTTAAGTTTCCCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCTCCGGGAGAGGAG 60
Db 5 ATTTAAGTTTCCCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCTCTGGGAGAGGAG 64
Qy 61 CGGCAACCGCGCGCGCCAG-CCCCAGCCCCATGAGACTCCGAGCAGGAGACTCGTGGGG 119
Db 65 CGCGCCCGCGCGCGCCCGCCCGCCAGCCCATGGACTCCGAGCAGGAGACTCGTGGGG 124
Qy 120 ATGTTAGCGTGTGCTGACAGGTGCTGCGACCTCCCTGCGAGTCCAGTGCAGCTCTCAATCGC 179
Db 125 ATGTTAGCTTGGCTATGACAGGTGCTGTGGACCTCCCTGCGAGTGCAGCTCTTAATCGC 184
Qy 180 ACAGGGGACCCAGGCGCTGGCCCTCCATCCAGAAAACTATGACCTCAACCGCTACCTG 239
Db 185 ACAGGAGATCCAGGCGCTGGCCCTCCATCCAGAAAACTATGACCTCAACCGCTACCTG 244
Qy 240 GAGCAACAACTCCGAGCTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAAC 299
Db 245 GAGCATCAACTCCGAGCTTAGCTGGGACTTACCTGAACTACTCTGGGGCCCCCTTTCAAC 304
Qy 300 GAGCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGAC 359
Db 305 GAGCCTGACTTCAATCTCTCGACTGGGGGAGAAACTCTGCCAGGGCCACGCTCAAC 364
Qy 360 TTGGAGGTGTGGGAAAGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCGTAC 419
Db 365 TTGGAAGTGTGGGAAAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGCGTAC 424
Qy 420 AGCAACCTTCTGTGTTACTTGGTGGGCTCAACCGTCAAGCTGCGCTGCTGAGCTGCGC 479
Db 425 AGTCACCTCTGTGTTACTTGGTGGGCTCAACCGTCAAGCTGCGCTGCGCTGAACTCCGA 484
Qy 480 CGAGCTTGGCGGCTTCTGACAGGCTTCCAGGGGCTCTGCGGAGAGCTTGGGGGCTC 539
Db 485 CGTAGCTTGGCGGCTTCTGTACAGGCTTCCAGGGGCTCTGCGGAGAGCTTGGAGGTTC 544
Qy 540 ATGGCAGCTTGGGCTACCCACTGCGCCAGCGGCTGCTGGGACTGAACCCACTTGGACT 599
Db 545 ATGGCGAGCTTGGCTTACCCACTGCGCCAGCGCTTCTGCCAGGAGCTGAGCCAGCTGGGC 604
Qy 600 CTGGCCCTGCGCCACAGTGACTTCTCCAGAAAGATGGAGACTTCTGGCTGCTGAAGGAG 659
Db 605 CTGGCCCTGCGCCACAGTGACTTCTCCAGAAAGATGGAGACTTCTGGCTGCTGAAGGAG 664
Qy 660 CTGCAAGCTTGGCTGCGGCTCGGCAAGGACTTCAACCGGCTCAAGAAAGATGTCAG 719
Db 665 CTGCAAGCTTGGCTATGGGTTACGCAAGGACTTCAACCGGCTTAAAGAAAGATGTCAG 724
Qy 720 CTCCAGAGCTGAGTCAACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACCTT 778
Db 725 CTCCAGAGCTTCACTCACTGACCTTGGAGGACATGGTTTCTGACCTCTGACCT 783
```

## RESULT 3

US-09-931-704-3

; Sequence 3, Application US/09931704

; Patent No. US20020041873A1

; GENERAL INFORMATION:

; APPLICANT: Senaldi, Giorgio

; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using

; TITLE OF INVENTION: Inhibitors

```
; FILE REFERENCE: A-695
; CURRENT APPLICATION NUMBER: US/09/931.704
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 5087
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (137)..(138)
; OTHER INFORMATION: product = "INTERVENING UNSEQUENCED REGION OF >1KB"
US-09-931-704-3

Query Match 65.7%; Score 523.4; DB 10; Length 5087;
Best Local Similarity 99.8%; Pred. No. 1e-115;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 271 ATCTGAACCTACTGGGCCCCCTTTCAAGAGCCAGACTTCAACCCCTCCCGCCTGGGGG 330
Db 363 AGCTGAACCTACTGGGCCCCCTTTCAAGAGCCAGACTTCAACCCCTCCCGCCTGGGGG 3422
Qy 331 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 390
Db 3423 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 3482
Qy 391 TCGCGCTGACCCAGAACTACAGAGGCTACAGCCACTTCTGTGTACTTGTGGCTGCA 450
Db 3483 TCGCGCTGACCCAGAACTACAGAGGCTACAGCCACTTCTGTGTACTTGTGGCTGCA 3542
Qy 451 ACCGTCAGGCTGCCACTGCTGAGCTGCGCGCAGCCTGGCCCACTTCTGCACGACCTCC 510
Db 3543 ACCGTCAGGCTGCCACTGCTGAGCTGCGCGCAGCCTGGCCCACTTCTGCACGACCTCC 3602
Qy 511 AGGGCTCTGTGGGAGCATTTGGGGCGTCAATGGCAGCTCTGGGCTACCCACTGCCCCAGC 570
Db 3603 AGGGCTCTGTGGGAGCATTTGGGGCGTCAATGGCAGCTCTGGGCTACCCACTGCCCCAGC 3662
Qy 571 CGCTCCCTGGGACTGAACCCACTTGGACTCTGGGCTGCGCCACAGTGAATCTCTCCAGA 630
Db 3663 CGCTCCCTGGGACTGAACCCACTTGGACTCTGGGCTGCGCCACAGTGAATCTCTCCAGA 3722
Qy 631 AGATGAGCAGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCTGGCCAAAG 690
Db 3723 AGATGAGCAGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCTGGCCAAAG 3782
Qy 691 ACTTCAACCGGCTCAAGAAGATGCAAGCTCCAGCAGCTGCAGTGCAGTCAACCTGCACCTGG 750
Db 3783 ACTTCAACCGGCTCAAGAAGATGCAAGCTCCAGCAGCTGCAGTGCAGTCAACCTGCACCTGG 3842
Qy 751 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 795
Db 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3887

RESULT 4
US-09-864-761-23175/c
; Sequence 23175, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
```

```
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23175
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUATE 8.00e-03
; OTHER INFORMATION: NT HIT: g11439486, EVALUATE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: A1752561.1, EVALUATE 0.00e+00
US-09-864-761-23175

Query Match 61.9%; Score 493.4; DB 10; Length 495;
Best Local Similarity 99.8%; Pred. No. 8e-109;
Matches 494; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 273 CTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGGGCA 332
Db 495 CTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGGGCA 436
Qy 333 GAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAACTG 392
Db 435 AAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAACTG 376
Qy 393 CGGTGACCCAGAACTACAGAGCCTTACAGCCACTTCTGTGTACTTGTGGTGGCTTCAAC 452
Db 375 CGGTGACCCAGAACTACAGAGCCTTACAGCCACTTCTGTGTACTTGTGGTGGCTTCAAC 316
Qy 453 CGTCAGGCTGCCACTGCTGAGCTGGCGCCGAGCCTTCTTCTGACCCAGCCTTCTGACCCAGCCTCCAG 512
```

|||||  
Db 315 CGTCAGGCTGCCACTGTAGCTGGCGCGAGCGTGGCCCACTTCTGCACGAGCTCCAG 256  
Qy 513 GGCCTCTGGGAGCATTTGGGGCTCATGCGAGCTCTGGGCTACCCACTGGCCCGCCAGCCG 572  
Db 255 GGCCTCTGGGAGCATTTGGGGCTCATGCGAGCTCTGGGCTACCCACTGGCCCGCCAGCCG 196  
Qy 573 CTGCTGGGACTGAACCACTTGGACTCTCTGGGCTCCGACAGTGCATCTCTCCAGAG 632  
Db 195 CTGCTGGGACTGAACCACTTGGACTCTCTGGGCTCCGACAGTGCATCTCTCCAGAG 136  
Qy 633 ATGAGCACTTTGGCTGTGTAAGGAGTGCAGACCTTGGCTGTGGGCTCGGCCCAAGGAC 692  
Db 135 ATGAGCACTTTGGCTGTGTAAGGAGTGCAGACCTTGGCTGTGGGCTCGGCCCAAGGAC 76  
Qy 693 TTCAACCGCTCAAGAGAGATGCAGCTCCAGCAGTGCATCTCCAGCTGCACCTGGG 752  
Db 75 TTCAACCGCTCAAGAGAGATGCAGCTCCAGCAGTGCATCTCCAGCTGCACCTGGG 16  
Qy 753 GCTCATGGCTTCTGA 767  
Db 15 GCTCATGGCTTCTGA 1

RESULT 5  
US-09-864-761-6462/c  
; Sequence 6462, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeonica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 6462  
; LENGTH: 492  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005849.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5  
; US-09-864-761-6462  
Query Match 46.3%; Score 369; DB 10; Length 492;  
Best Local Similarity 100.0%; Pred. No. 3.3e-79;  
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 TTCTGTGTACTTGGTGGCTTCAACCGTCAAGGCTGCGCTGCTGAGCTGCGCCGCGAGCC 486  
Db 492 TTCTGTGTACTTGGTGGCTTCAACCGTCAAGGCTGCGCTGCTGAGCTGCGCCGCGAGCC 433  
Qy 487 TGGCCCACTTCTGCACAGCTCCAGGCGCTGCTGGGAGCATTTGGGGCTCATGGCAG 546  
Db 432 TGGCCCACTTCTGCACAGCTCCAGGCGCTGCTGGGAGCATTTGGGGCTCATGGCAG 373  
Qy 547 CTCTGGGCTACCCACTGCGCCCGAGCGCTGCTGGGAGTGAACCCACTTGGACTTCCTGGCC 606  
Db 372 CTCTGGGCTACCCACTGCGCCCGAGCGCTGCTGGGAGTGAACCCACTTGGACTTCCTGGCC 313  
Qy 607 CTGCCCCAGTGACTTCTCCAGAGATGGAGCACTTCTGGCTGTGAAGGAGCTGCAGA 666  
Db 312 CTGCCCCAGTGACTTCTCCAGAGATGGAGCACTTCTGGCTGTGAAGGAGCTGCAGA 253  
Qy 667 CTGGCTGTGGGCTCGGCCAGGACTTCAACCGCTCAAGAGATGAGCGCTCCAG 726  
Db 252 CTGGCTGTGGGCTCGGCCAGGACTTCAACCGCTCAAGAGATGAGCGCTCCAG 193  
Qy 727 CAGCTGCAGTCAACCTGCACTGGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTT 786  
Db 192 CAGCTGCAGTCAACCTGCACTGGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTT 133  
Qy 787 CGCTCCCCC 795  
Db 132 CGCTCCCCC 124

RESULT 6  
US-09-765-231A-75  
; Sequence 75, Application US/09765231A  
; Patent No. US20020119452A1  
; GENERAL INFORMATION:  
; APPLICANT: Searle/Monsanto  
; APPLICANT: Phippard, Deborah  
; APPLICANT: Vasanthakamur, Geetha  
; APPLICANT: Dotson, Stanton  
; APPLICANT: Ma, Xiao-Jun  
; TITLE OF INVENTION: Osteoarthritis tissue-derived nucleic acids, polypeptides,  
; TITLE OF INVENTION: vectors, and cells  
; FILE REFERENCE: SO-3221 ER  
; CURRENT APPLICATION NUMBER: US/09/765,231A  
; CURRENT FILING DATE: 2001-01-18  
; NUMBER OF SEQ ID NOS: 82  
; SEQ ID NO 75  
; LENGTH: 283  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-765-231A-75



Query Match 33.8%; Score 269.4; DB 10; Length 283;  
Best Local Similarity 99.3%; Pred. No. 1.5e-55;  
Matches 281; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
  
QY 37 CTCGCCAGCTCCGGGAGAGAGCCGACCGCGCGCGCCAGCCCGCCAGCCCGCCATGGACC 96  
|||  
DB 1 CTCGCCAGCTCCGGGAGAGAGCCGACCGCGCGCGCCAGCCCGCCAGCCCGCCATGGACC 60  
|||  
  
QY 97 TCCGAGCAGGGGACT- CCGTGGGGATGTTAGCGTCCTGTGACCGGTGCTTGGCACCTC 155  
|||  
DB 61 TCCGAGCAGGGGACTCGCTGGGGATGTTAGCGTCCTGTGACCGGTGCTTGGCACCTC 120  
|||  
  
QY 156 CTGAGTGCAGCTCTCAATCGACAGGGGACCCAGGCGCTGGCCCTTCCATCCAGAAA 215  
|||  
DB 121 CTGAGTGCAGCTCTCAATCGACAGGGGACCCAGGCGCTGGCCCTTCCATCCAGAAA 180  
|||  
  
QY 216 ACCTATGACCTCACCCGCTACTCTGAGGACCAACTCCGAGCTTGGCTGGGACCTATCTG 275  
|||  
DB 181 ACCTATGACCTCACCCGCTACTCTGAGGACCAACTCCGAGCTTGGCTGGGACCTATCTG 240  
|||  
  
QY 276 AACTACTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTC 318  
|||  
DB 241 AACTACTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTC 283  
|||  
  
RESULT 7  
US-10-014-717-1  
; Sequence 1, Application US/10014717  
; Publication No. US20020192778A1  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goriach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/10/014,717  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US/09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-10-014-717-1  
  
Query Match 5.6%; Score 44.8; DB 9; Length 68750;  
Best Local Similarity 50.0%; Pred. No. 0.16;  
Matches 112; Conservative 0; Mismatches 112; Indels 0; Gaps 0;  
  
QY 456 CAGGTGCCATCTGAGCTGGCGGAGCTGGCCCACTTGTGACCAAGCTCCAGGCG 515  
|||  
DB 57992 CAGGTAGCGAGCTGGGAGCCGGGGAACGTGCGCGCCCTTCCGATCCACGCGCGGC 58051  
|||  
  
QY 516 CTGCTGGGACATTCGGGGGTATGCGAGCTTGGGCTACCCACTGCCCGCAGCGCTG 575  
|||  
DB 58052 GTGCTGGCGCGGATCGGGCGGTGAGACCCCGCGAGGTGCAACCTCGCGCCCTGCGC 58111  
|||  
  
QY 576 CTGGGACTGAACCACTTGGACTCTTGGCCCTGCGCCAGTGAATCTTCCACAGAGATG 635  
|||  
DB 58112 GCCCGGCTTATGCGCGCTGCGCGCTATCTATGTGTGCTGCGGAGATGGG 58171  
|||  
  
QY 636 GACGACTTCTGGCTGCTGAAGAGCTGCAGACCTGGCTGTGGCG 679  
|||  
DB 58172 CTTCAATACGCGCGCGCTTGGGGGGCTCGCGAGCTGTGGCG 58215  
|||  
  
RESULT 8

US-09-864-761-2889/c  
; Sequence 2889, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aesomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 2889  
; LENGTH: 460  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC007200.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.4  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 12  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.4  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.7  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.8  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.1  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.9  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.5  
US-09-864-761-2889  
  
Query Match 4.9%; Score 38.8; DB 10; Length 460;  
Best Local Similarity 58.8%; Pred. No. 1.3;  
Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 6 AGCTTCGCGGAGCGCGCTCGCTCCACTCCGCGGCTCCGCGAGGAGGCGCA 65  
Db 380 AGCTTCGCGGAGCGCGCTCGCTCCACTCCGCGGCTCCGCGAGGAGGCGCA 321  
QY 66 CCG 119  
Db 320 CCGTGTCCACGCGCACCGGAGTCCGCGCGCGCGCGCGCGCGCGCG 267

## RESULT 9

US-09-925-301-363  
; Sequence 363, Application US/09925301  
; Patent No. US2002052308A1

## GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA106

; CURRENT APPLICATION NUMBER: US/09/925,301

; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05882

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 1694

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 363

; LENGTH: 1324

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: misc feature

; LOCATION: (385)

; OTHER INFORMATION: n equals a,t,g, or c

US-09-925-301-363

Query Match 4.8%; Score 38.4; DB 10; Length 1324;  
Best Local Similarity 50.0%; Pred. No. 2.1;  
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 512 GGGCTGTGGGAGGAGTTCGGCGGCTGATGGAGCTCTGGGCTACCGCTCCCGAGCC 571  
Db 467 GGACAAGCTGGCGCGCTGCGAGTTCGACTCGCGGAGTATGGCTGCTCAAGGCGCATCGC 526  
QY 572 GCTGCTGGGAGTGAACCACTTGGACTCTCTGGCGCTGCGGAGTCTCTCCAGAA 631  
Db 527 GCTCTTACGCGCGAGCGCTGTGGCTCTCAGACCGCGCGCGCTTGGAGCGCTGCGAG 586  
QY 632 GATGAGAGTCTTCTGGCTGTGAAGGAGCTGCAGACCTGTGGGCTCGGCGCAAGGA 691  
Db 587 GAAGCGCAGGTGGCGCTACCGAGTATGTGGGCGCAGTACCGCTCCCGAGCGCG 646  
QY 692 CTTCAACCGGCT 703  
Db 647 CTTGGGCGGCT 658

## RESULT 10

US-09-764-868-1440/c  
; Sequence 1440, Application US/09764868  
; Patent No. US2002016871A1

## GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT232

; CURRENT APPLICATION NUMBER: US/09/764,868

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1510

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1440

; LENGTH: 11336

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (4205)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-868-1440

Query Match 4.8%; Score 38; DB 9; Length 11336;

Best Local Similarity 67.9%; Pred. No. 4.4;

Matches 53; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 412 AGGCTACAGCCACCTTCTGTGTACTTGGCTCAACCGTCAGGCTGCCACTGTCTG 471  
Db 1466 AGGCTCGGGACATCTGCAGTGGTCCAGCTGGCCCCCACCACAGCAGCCACTGCTG 1407

QY 472 AGCTGCGCGCGAGCTGG 489

Db 1406 TCTGCCCCATACTGG 1389

## RESULT 11

US-09-925-697-1/c

; Sequence 1, Application US/09925697

; Patent No. US20020119511A1

; GENERAL INFORMATION:

; APPLICANT: May, Earl

; APPLICANT: Van Horn, Stephanie

; APPLICANT: Warren, Patrick V.

; APPLICANT: Warren, Richard L.

; TITLE OF INVENTION: dnaE

; FILE REFERENCE: GM10237

; CURRENT APPLICATION NUMBER: US/09/925,697

; CURRENT FILING DATE: 2001-08-09

; PRIOR APPLICATION NUMBER: US/09/387,695

; PRIOR FILING DATE: 1999-08-31

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Fast-SEQ for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 3129

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-09-925-697-1

Query Match 4.7%; Score 37.8; DB 10; Length 3129;  
Best Local Similarity 47.3%; Pred. No. 3.7;  
Matches 114; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 479 CGCGAGCTGGCGGCACTTCGACCACTCCAGGCGCTGCGGAGGATTCGCGGCGT 538  
Db 2056 CTGCAAACTTCTCCATAACATCAAAACCTGCTCTGATTTTCCGAGTATGGCGCTT 1997  
QY 539 CATGCGAGCTGGGCTACCCACTGCCCGCGCTGCTGGGACTGAACCCACTTTGGAC 598  
Db 1996 CTATGGAGCTTGAATAAAGGAGCGCTCATCTCATGATAGCAGAGGATCTTTTAC 1937  
QY 599 TCCTGGCGCTGCCACAGTGAATCTCTCCAGAAGATGGAAGACTTCTGGCTGTGAAGGA 658  
Db 1936 CCATAGCTCGAGGCAAAATATCGGCTTCCCAAGACTAAATCCAGCAATCGCTGAGCAA 1877  
QY 659 GCTGAGAGCTGGCTGTGGCGCTCGGCGGAGGACTTCAACCGGCTCAAGNAGAGATGCA 718  
Db 1876 CTTGCATAACCTGCTCTCTGATAGAGCATAATGCCAATAGTTGGAGGCAAAATATCTCTCA 1817  
QY 719 G 719  
Db 1816 G 1816

## RESULT 12

US-09-954-456-45

; Sequence 45, Application US/09954456

; Patent No. US20020115057A1

; GENERAL INFORMATION:

; APPLICANT: Young, Paul

```

RESULT 13
US-09-954-456-1621
; Sequence 1621, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25

```

	Query Match	Best Local Similarity	52.8%	Score 37.4;	DB 10;	Length 3089;
	Matches 103;	Conservative	0;	Mismatches 91;	Indels	1; Gaps 1;
QY	135	TGCACGGTGTCTGGCACCTCCCTGCAGTCCAGCTCTCAATCGCACAGGGGACCCAGGG	194			
Db	407	TGCACAGACCTTGGGCCCCCATCTTCAGAAATCACTCCCTGCGGTGCTGCACAGGCCACGG	466			
QY	195	CTTGGCCCTTCCATCCAGAAAACTATGACCTACACCCGGTACTCTGGAGCACCAACTCCGC	254			

Db 467 CCAGCAGGATCC-CCTGGCAGACTACCACTGACAGCTGGATGGAGGAGGATACCGC 525  
QY 255 AGCTTGGCTGGGACCTATCTGAATCTGCTGGCCCCCTTTCAACGAGCCAGACTTCAAC 314  
Db 526 CACTTCCCGGGGAGTACGTCCTTACCGGCTACCCCTGCTCCTGGACCTTCTACCCAC 585  
QY 315 CCTCCCCGCTGGG 329  
Db 586 CACCTCCGCCAGGAG 600

RESULT 15  
US-09-954-531-151/C  
; Sequence 151, Application US/09954531  
; Patent No. US20020165180A1  
; GENERAL INFORMATION:  
; APPLICANT: Weaver, Zoe  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc  
; FILE REFERENCE: 689290-77  
; CURRENT APPLICATION NUMBER: US/09/954,531  
; CURRENT FILING DATE: 2002-05-02  
; PRIOR APPLICATION NUMBER: US/60/233,133  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,009  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,034  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,509  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: US/60/234,567  
; PRIOR FILING DATE: 2000-09-22  
; NUMBER OF SEQ ID NOS: 1392  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 151  
; LENGTH: 3984  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-531-151

Query Match 4.7%; Score 37.2; DB 9; Length 3984;  
Best Local Similarity 52.6%; Pred. No. 5.4;  
Matches 81; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 560 ACTGCCCCAGCCGCTGCTGGGACTGACCCACTTGGACTCTGGCCCTGCCACAGTGA 619  
Db 1081 ACTGGCCTCTGCGCATGCTCTCCGGCAGCCACTCGGCCCTCTGGCTATGCTCTCTGGG 1022  
QY 620 CTTCTCCAGAGATGGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCG 679  
Db 1021 GACCACCTGCATGAGCTTCTGACACAGGACGCTCGTGAGAGATGCCCTGGAGGAGCCA 962  
QY 680 CTCGGCCAAGACTTCAACCGGCTCAAGAGAAG 713  
Db 961 CTTGGGCTTGTTTTCCACCAGACCCCGAGAAG 928

Search completed: February 2, 2003, 03:55:32  
Job time : 102.244 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 25.6828 Seconds  
(without alignments)  
9516.918 Million cell updates/sec

Title: US-09-931-704-1  
Perfect score: 797  
Sequence: 1 attaaagcttcgcccagacc.....tctctcttctgctccccccc 797

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	797	100.0	797	1	US-08-792-019B-1	Sequence 1, Appli
2	797	100.0	797	3	US-08-988-819-1	Sequence 1, Appli
3	797	100.0	797	3	US-09-016-534-1	Sequence 1, Appli
4	749.4	94.0	1710	3	US-09-106-182-1	Sequence 1, Appli
5	669.4	84.0	819	1	US-08-792-019B-4	Sequence 4, Appli
6	669.4	84.0	819	3	US-08-988-819-4	Sequence 4, Appli
7	669.4	84.0	819	3	US-09-016-534-4	Sequence 4, Appli
8	523.4	65.7	5087	1	US-08-792-019B-3	Sequence 3, Appli
9	523.4	65.7	5087	3	US-08-988-819-3	Sequence 3, Appli
10	523.4	65.7	5087	3	US-09-016-534-3	Sequence 3, Appli
11	80.2	10.1	396	3	US-09-106-182-7	Sequence 7, Appli
12	44.8	5.6	68750	3	US-09-335-409-1	Sequence 1, Appli
13	44.8	5.6	68750	4	US-09-568-102-1	Sequence 1, Appli
14	44.8	5.6	68750	4	US-09-567-969-1	Sequence 1, Appli
15	44.8	5.6	68750	4	US-09-568-480-1	Sequence 1, Appli
16	44.8	5.6	68750	4	US-09-568-486-1	Sequence 1, Appli
17	44.8	5.6	68750	4	US-09-568-472-1	Sequence 1, Appli
18	44.8	5.6	68750	4	US-09-567-899-1	Sequence 1, Appli
19	38.4	4.8	71989	4	US-09-443-501A-2	Sequence 2, Appli
20	38	4.8	223	1	US-08-383-761-1	Sequence 1, Appli
21	38	4.8	223	1	US-08-824-277-1	Sequence 1, Appli
22	37.8	4.7	3129	4	US-09-387-695-1	Sequence 1, Appli
23	37.8	4.7	43272	1	US-08-614-770A-1	Sequence 1, Appli
24	36.8	4.6	1308	4	US-08-937-067-18	Sequence 18, Appli
25	36.8	4.6	4695	2	US-08-231-193A-57	Sequence 57, Appli
26	36.8	4.6	4695	2	US-08-486-273A-57	Sequence 57, Appli
27	36.8	4.6	4695	3	US-08-940-086A-57	Sequence 57, Appli

c	28	36.8	4.6	4695	4	US-08-940-035A-57	Sequence 57, Appli
c	29	36.8	4.6	4695	4	US-08-935-105A-57	Sequence 57, Appli
c	30	36.8	4.6	4695	4	US-09-648-797-57	Sequence 57, Appli
	31	36.4	4.6	645	1	US-08-312-870-4	Sequence 4, Appli
	32	36.4	4.6	1338	1	US-08-307-444A-8	Sequence 8, Appli
	33	36.4	4.6	1338	1	US-08-587-389-8	Sequence 8, Appli
	34	36.4	4.6	1368	1	US-08-307-444A-7	Sequence 7, Appli
	35	36.4	4.6	1368	1	US-08-587-389-7	Sequence 7, Appli
	36	36.4	4.6	1425	1	US-08-307-444A-6	Sequence 6, Appli
	37	36.4	4.6	1425	1	US-08-307-444A-23	Sequence 23, Appli
	38	36.4	4.6	1425	1	US-08-587-389-6	Sequence 6, Appli
	39	36.4	4.6	1425	1	US-08-587-389-23	Sequence 23, Appli
	40	36.4	4.6	1428	1	US-08-014-723-3	Sequence 3, Appli
	41	36.4	4.6	1428	1	US-08-014-723-4	Sequence 4, Appli
	42	36.4	4.6	1428	1	US-08-110-011A-3	Sequence 3, Appli
	43	36.4	4.6	1428	1	US-08-110-011A-4	Sequence 4, Appli
	44	36.4	4.6	1491	1	US-08-312-870-2	Sequence 2, Appli
	45	36.4	4.6	1680	1	US-08-014-723-13	Sequence 13, Appli

ALIGNMENTS

RESULT 1  
US-08-792-019B-1  
; Sequence 1, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUTROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/792,019B  
; FILING DATE: 03-FEB-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 797 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 90..764  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 171..764  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 90..170  
; US-08-792-019B-1

Query Match 100.0%; Score 797; DB 1; Length 797;  
Best Local Similarity 100.0%; Pred. No. 3.2e-161;  
Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATTAAGCTTCGCGAGCGCGCTCGCCCTCCACTCCGCGAGCTCCGCGAGGAG 60
Db 1 ATTAAGCTTCGCGAGCGCGCTCGCCCTCCACTCCGCGAGCTCCGCGAGGAG 60
QY 61 CCGCACC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 CCGCACC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 TGTAGCGTCTGTCACCGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 121 TGTAGCGTCTGTCACCGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 181 CAGGGAGCCAGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 CAGGGAGCCAGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 241 AGACCAACTCCGCGAGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 241 AGACCAACTCCGCGAGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 301 AGCAGACTTCAACCCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 301 AGCAGACTTCAACCCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 361 TGGAGGTGTGCGAAGCTCAATGACAACTGCGCGCTGACCGAGAACTACGAG 420
Db 361 TGGAGGTGTGCGAAGCTCAATGACAACTGCGCGCTGACCGAGAACTACGAG 420
QY 421 GCACCTTCTGTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Db 421 GCACCTTCTGTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 481 GCAGCTTCTGTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 481 GCAGCTTCTGTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
QY 541 TGGCAGCTTGGGTACCCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db 541 TGGCAGCTTGGGTACCCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
QY 601 CTGCGCTGCGCGAGTACTTCTCCAGAGATGGAGACTTCTGGCTGTGAAGGAG 660
Db 601 CTGCGCTGCGCGAGTACTTCTCCAGAGATGGAGACTTCTGGCTGTGAAGGAG 660
QY 661 TGCAGACTTGGCTGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Db 661 TGCAGACTTGGCTGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 721 CTCAGAGCTGCGAGTCAACCTGACCTGGGGGCTCATGGCTTCTGACTTCTG 780
Db 721 CTCAGAGCTGCGAGTCAACCTGACCTGGGGGCTCATGGCTTCTGACTTCTG 780
QY 781 CCTCTTCTGCTCCCCCCC 797
Db 781 CCTCTTCTGCTCCCCCCC 797
```

## RESULT 2

```
US-08-988-819-1
; Sequence 1, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..764
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 171..764
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..170
; US-08-988-819-1
```

Query Match 100.0%; Score 797; DB 3; Length 797;

Best Local Similarity 100.0%; Pred. No. 3.2e-181; Indels 0; Gaps 0;

Matches 797; Conservative 0; Mismatches 0;

```
QY 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCACTCCGCGAGCTCCGCGAGGAG 60
Db 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCACTCCGCGAGCTCCGCGAGGAG 60
QY 61 CCGCACC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 CCGCACC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 TGTAGCGTCTGTCACCGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 121 TGTAGCGTCTGTCACCGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 181 CAGGGAGCCAGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 CAGGGAGCCAGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 241 AGCACAACCTCCGCGAGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 241 AGCACAACCTCCGCGAGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 301 AGCCAGACTTCAACCCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 301 AGCCAGACTTCAACCCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 361 TGGAGGTGTGCGAAGCTCAATGACAACTGCGCGCTGACCGAGAACTACGAG 420
Db 361 TGGAGGTGTGCGAAGCTCAATGACAACTGCGCGCTGACCGAGAACTACGAG 420
QY 421 GCACCTTCTGTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Db 421 GCACCTTCTGTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 481 GCAGCTTCTGTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 481 GCAGCTTCTGTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
```

Qy 541' TGGCAGCTCTGGGCTACCCACTGCCCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 600  
Db 541 TGGCAGCTCTGGGCTACCCACTGCCCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 600  
Qy 601 CTGGCCCTGCCCCACAGTACTTCTCCAGAGATGGAGCTTCTGGCTGTGAAGGAGC 660  
Db 601 CTGGCCCTGCCCCACAGTACTTCTCCAGAGATGGAGCTTCTGGCTGTGAAGGAGC 660  
Qy 661 TGCAGACCTGGCTGTGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 720  
Db 661 TGCAGACCTGGCTGTGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 720  
Qy 721 CTCAGAGCTGCTAGTCACTGACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACTTCT 780  
Db 721 CTCAGAGCTGCTAGTCACTGACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACTTCT 780  
Qy 781 CCTCTTGGCTCCCCCCC 797  
Db 781 CCTCTTGGCTCCCCCCC 797

## RESULT 3

US-09-016-534-1  
; Sequence 1, Application US/09016534  
; Patent No. 6143874  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; APPLICANT: ELLIOTT, GARY S.  
; APPLICANT: SARMIENTO, ULLA  
; APPLICANT: SENALDI, GIORGIO  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,534  
; FILING DATE:

## CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442B  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 797 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 90..764  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 171..764  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 90..170  
; US-09-016-534-1

Query Match 100.0%; Score 797; DB 3; Length 797;  
Best Local Similarity 100.0%; Pred. No. 3.2e-181;  
Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATTTAAGCTTCCGCGAGCCGCGGCTCGCCCTCCCACTCCGCGAGCTTCCGGGAGAGAG 60  
Db 1 ATTTAAGCTTCCGCGAGCCGCGGCTCGCCCTCCCACTCCGCGAGCTTCCGGGAGAGAG 60  
Qy 61 CCGCACCCGCGCGGCCAGAGCCCAAGCCCATGGACCTCCGAGAGGGGACTCGTGGGGGA 120  
Db 61 CCGCACCCGCGCGGCCAGAGCCCAAGCCCATGGACCTCCGAGAGGGGACTCGTGGGGGA 120  
Qy 121 TGTAGCGTCTGTGACAGGTGTCTTGGCACCTCCCTGCGAGTGCAGCTCTCAATCGCA 180  
Db 121 TGTAGCGTCTGTGACAGGTGTCTTGGCACCTCCCTGCGAGTGCAGCTCTCAATCGCA 180  
Qy 181 CAGGGACCCAGGGCTGGCCCCCTCCATCCAGAAAACTATGACCTACCCGCTACCTGG 240  
Db 181 CAGGGACCCAGGGCTGGCCCCCTCCATCCAGAAAACTATGACCTACCCGCTACCTGG 240  
Qy 241 AGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTCTGGGCCCCCTTTCAACG 300  
Db 241 AGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTCTGGGCCCCCTTTCAACG 300  
Qy 301 AGCCAGACTTCAACCCCTCCCGCTGGGGGCGAGAGACTCTGCCAGGGCCACTGTGTACT 360  
Db 301 AGCCAGACTTCAACCCCTCCCGCTGGGGGCGAGAGACTCTGCCAGGGCCACTGTGTACT 360  
Qy 361 TGGAGGTGTGGGAAGCCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCTTACA 420  
Db 361 TGGAGGTGTGGGAAGCCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCTTACA 420  
Qy 421 GCCACCTTCTGTGTACTTGTGGTGGCCCTCAACCGTCAAGCTGCTGAGCTGCGCC 480  
Db 421 GCCACCTTCTGTGTACTTGTGGTGGCCCTCAACCGTCAAGCTGCTGAGCTGCGCC 480  
Qy 481 GCAGCTGGCCCCACTTCTGCACACGCTTCCAGGGCTCTCTGGGCGAGCAATTGGGGCGTCA 540  
Db 481 GCAGCTGGCCCCACTTCTGCACACGCTTCCAGGGCTCTCTGGGCGAGCAATTGGGGCGTCA 540  
Qy 541 TGGCAGCTCTGGGCTACCCACTGCCCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 600  
Db 541 TGGCAGCTCTGGGCTACCCACTGCCCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 600  
Qy 601 CTGGCCCTGCCCCACAGTACTTCTCCAGAGATGGAGCTTCTGGCTGTGAAGGAGC 660  
Db 601 CTGGCCCTGCCCCACAGTACTTCTCCAGAGATGGAGCTTCTGGCTGTGAAGGAGC 660  
Qy 661 TGCAGACCTGGCTGTGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 720  
Db 661 TGCAGACCTGGCTGTGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 720  
Qy 721 CTCAGAGCTGCTAGTCACTGACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACTTCT 780  
Db 721 CTCAGAGCTGCTAGTCACTGACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACTTCT 780  
Qy 781 CCTCTTGGCTCCCCCCC 797  
Db 781 CCTCTTGGCTCCCCCCC 797

## RESULT 4

US-09-106-182-1  
; Sequence 1, Application US/09106182  
; Patent No. 6046035  
; GENERAL INFORMATION:  
; APPLICANT: Shi, Yanggu  
; APPLICANT: Ruben, Steve  
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc  
; STREET: 9410 Key West Ave

QY 465 ACTGCTGAGCTGCGCCGAGCGCTGCGCCACCTTCTGCAACAGCGCTCCAGGGCTGCTGGGC 524  
Db 421 ACTGCTGAGCTGCGCCGAGCGCTGCGCCACCTTCTGCAACAGCGCTCCAGGGCTGCTGGGC 480  
QY 525 AGCATTGGGGGGTTCATGGCAGCTTGGGCTACCCAGCTGCGCCAGCGCTGCTGGGACT 584  
Db 481 AGCATTGGGGGGTTCATGGCAGCTTGGGCTACCCAGCTGCGCCAGCGCTGCTGGGACT 540  
QY 585 GAAACCCACTTGGACTTCTGCGCCCTGCCACAGTGAATCTCTCCAGAAAGATGGAACGACTTC 644  
Db 541 GAAACCCACTTGGACTTCTGCGCCCTGCCACAGTGAATCTCTCCAGAAAGATGGAACGACTTC 600  
QY 645 TGGCTGTGAAGAGAGCTGCAGACCTGGGCTGTGGCGTGGCGTCCAGAGACTTCAACCGGCTC 704  
Db 601 TGGCTGTGAAGAGAGCTGCAGACCTGGGCTGTGGCGTGGCGTCCAGAGACTTCAACCGGCTC 660  
QY 705 AAGAAGAAGATGCAGCTTCCAGCAGCTGCAGTCAACCTGCACTGGGGCTCATGGCTTC 764  
Db 661 AAGAAGAAGATGCAGCTTCCAGCAGCTGCAGTCAACCTGCACTGGGGCTCATGGCTTC 720  
QY 765 TGACTTCTGACCTTCT 795  
Db 721 TGACTTCTGACCTTCT 751

RESULT 5  
US-08-792-019B-4  
; Sequence 4, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/792,019B  
; FILING DATE: 03-FEB-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 819 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 95..769  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 176..769  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 95..175  
; US-08-792-019B-4

Query Match 84.0%; Score 669.4; DB 1; Length 819;  
Best Local Similarity 92.0%; Pred. No. 7.7e-151;

QY 45 GCCTCGGGAGAGAGCGCCAGCCGCGCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCA 104  
Db 1 GCCTCGGGAGAGAGCGCCAGCCGCGCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCA 60  
QY 105 GGGGACTCGTGGGGGATGTTAGCGTGCTGTGCAAGGTGCTCTGGCACCTCCCTGCACTG 164  
Db 61 GGGGACTCGTGGGGGATGTTAGCGTGCTGTGCAAGGTGCTCTGGCACCTCCCTGCACTG 120  
QY 165 CAGCTCTCAATGCACAGGGGACCCAGGGGCTGGGCGCTTCATCCAGAAACCTATGAC 224  
Db 121 CAGCTCTCAATGCACAGGGGACCCAGGGGCTGGGCGCTTCATCCAGAAACCTATGAC 180  
QY 225 CTCACCGCTACTGGAGACCACTCCGAGCTTGGCTGGGACTATCTCAACTACCTG 284  
Db 181 CTCACCGCTACTGGAGACCACTCCGAGCTTGGCTGGGACTATCTCAACTACCTG 240  
QY 285 GGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGGGGAGAGACTCTGCCC 344  
Db 241 GGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGGGGAGAGACTCTGCCC 300  
QY 345 AGGCCACCTTGAATGGAGGTGGGAGAGCCCTCAATGACAAACTCGCGCTGACCCAG 404  
Db 301 AGGCCACCTTGAATGGAGGTGGGAGAGCCCTCAATGACAAACTCGCGCTGACCCAG 360  
QY 405 AACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGGCCCTCAACCGTCAGCGTCC 464  
Db 361 AACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGGCCCTCAACCGTCAGCGTCC 420

US-09-106-182-1  
Query Match 94.0%; Score 749.4; DB 3; Length 1710;  
Best Local Similarity 99.9%; Pred. No. 8.1e-170;  
Matches 750; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 45 GCCTCGGGAGAGAGCGCCAGCCGCGCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCA 104  
Db 1 GCCTCGGGAGAGAGCGCCAGCCGCGCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCA 60  
QY 105 GGGGACTCGTGGGGGATGTTAGCGTGCTGTGCAAGGTGCTCTGGCACCTCCCTGCACTG 164  
Db 61 GGGGACTCGTGGGGGATGTTAGCGTGCTGTGCAAGGTGCTCTGGCACCTCCCTGCACTG 120  
QY 165 CAGCTCTCAATGCACAGGGGACCCAGGGGCTGGGCGCTTCATCCAGAAACCTATGAC 224  
Db 121 CAGCTCTCAATGCACAGGGGACCCAGGGGCTGGGCGCTTCATCCAGAAACCTATGAC 180  
QY 225 CTCACCGCTACTGGAGACCACTCCGAGCTTGGCTGGGACTATCTCAACTACCTG 284  
Db 181 CTCACCGCTACTGGAGACCACTCCGAGCTTGGCTGGGACTATCTCAACTACCTG 240  
QY 285 GGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGGGGAGAGACTCTGCCC 344  
Db 241 GGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGGGGAGAGACTCTGCCC 300  
QY 345 AGGCCACCTTGAATGGAGGTGGGAGAGCCCTCAATGACAAACTCGCGCTGACCCAG 404  
Db 301 AGGCCACCTTGAATGGAGGTGGGAGAGCCCTCAATGACAAACTCGCGCTGACCCAG 360  
QY 405 AACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGGCCCTCAACCGTCAGCGTCC 464  
Db 361 AACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGGCCCTCAACCGTCAGCGTCC 420



```
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
QY 1 ATTAAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCCAGCTCTCGGAGAGGAG 60
Db 1 ATTAAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCCAGCTCTCGGAGAGGAG 64
QY 61 CGCAGCCGCGCGCGCCAG-CCCCAGCCCATGAGCTCCGAGCAGGAGCTCTGTTGGGG 119
Db 65 CGCAGCCGCGCGCGCGCCAG-CCCCAGCCCATGAGCTCCGAGCAGGAGCTCTGTTGGGG 124
QY 120 ATGTTAGCTGCTGTGACAGCTGTCTGCGACCTCTCCCTCGAGTGCAGCTCTCAATCG 179
Db 125 ATGTTAGCTGCTGTGACAGCTGTCTGCGACCTCTCCCTCGAGTGCAGCTCTCAATCG 184
QY 180 ACAGGGGACCCAGGGCTCGCCCTCCATCCAGAAACCTATGACCTCACCGGCTACCTG 239
Db 185 ACAGGAGATCCAGGCTCGCCCTCCATCCAGAAACCTATGACCTCACCGGCTACCTG 244
QY 240 GAGCACCACCTCCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAAC 299
Db 245 GAGCATCACTCCGAGCTTAGCTGGGACCTACCTGAACTACCTGGGCCCCCTTTCAAC 304
QY 300 GAGCAGACTTCAACCTCCCGCTCGCGGCGGAGACTCTGCCAGGGCCACTGTTGAC 359
Db 305 GAGCCTGACTTCAATCTCTCGACTGGGGCGAGAACTCTGCCAGGGCCAGCTGTTGAC 364
QY 360 TTGGAGTGTGGCGAAGCTCAATGACAACTCGCGCTGACCCAGAACTACGAGGCTAC 419
Db 365 TTGGAGTGTGGCGAAGCTCAATGACAACTCGCGCTGACCCAGAACTATGAGGCTAC 424
QY 420 AGCCACTTCTGTGTACTGTTGGTGGGACCTATCTGAACTACTGGGCCCCCTTTCAAC 479
Db 425 AGTCACTCTCTGTGTACTGTTGGTGGGACCTCAACCGCTCAGGCTGCACTGAACTCCGA 484
QY 480 CGCAGCTGGGCCACTTCTGACAGCTCCAGGCTCGCGGCTGACCCAGAACTACGAGGCTC 539
Db 485 CGTACCTGGGCCACTTCTGACAGCTCCAGGCTCGCGGCTGACCCAGAACTATGAGGCTC 544
QY 540 ATGGAGCTCTGGGCTACCCACTGCGCCAGCGCTCGCTGGGACTGAACTTGGACT 599
Db 545 ATGGAGCTCTGGGCTACCCACTGCGCCAGCGCTCGCTGGGACTGAACTTGGACT 604
QY 600 CTGGCCCTGCGCCAGTACTTCTCCAGAGATGAGCACTTCTGGTGTCTGAAGGAG 659
Db 605 CTGGCCCTGCGCCAGTACTTCTCCAGAGATGAGCACTTCTGGTGTCTGAAGGAG 664
QY 660 CTGAGACTGCTGTGGGCTCGCGCAGGACTTCAACCGCTCAAGAGAGATGACG 719
Db 665 CTGAGACTGCTGTGGGCTCGCGCAGGACTTCAACCGCTCAAGAGAGATGACG 724
QY 720 CTTCCAGAGCTGAGTCACTCCCTGACCTGGGGGCTCTGAGTCTTCTGACCTT 778
Db 725 CTTCCAGAGCTTCACTCACTCCCTGACCTGGAGGACATGTTTCTGACCTTCACT 783
```

RESULT 6

```
US-08-988-819-4
; Sequence 4, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 819 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..769
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 176..769
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..175
; US-08-988-819-4
```

Query Match 84.0%; Score 669.4; DB 3; Length 819;

Best Local Similarity 92.0%; Pred. No. 7.7e-151;

Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

```
QY 1 ATTAAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCCAGCTCTCGGAGAGGAG 60
Db 5 ATTAAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCCAGCTCTCGGAGAGGAG 64
QY 61 CGCAGCCGCGCGCGCCAG-CCCCAGCCCATGAGACTCTGCCAGGAGGAGCTCTGTTGGGG 119
Db 65 CGCAGCCGCGCGCGCGCCAG-CCCCAGCCCATGAGACTCTGCCAGGAGGAGCTCTGTTGGGG 124
QY 120 ATGTTAGCTGCTGTGACAGCTGTCTGCGACCTCTCGGACCTCTCCCTCGAGTGCAGCTCTCAATCG 179
Db 125 ATGTTAGCTGCTGTGACAGCTGTCTGCGACCTCTCCCTCGAGTGCAGCTCTCAATCG 184
QY 180 ACAGGGGACCCAGGGCTCGCCCTCCATCCAGAAACCTATGACCTCACCGGCTACCTG 239
Db 185 ACAGGAGATCCAGGCTCGCCCTCCATCCAGAAACCTATGACCTCACCGGCTACCTG 244
QY 240 GAGCACCACCTCCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAAC 299
Db 245 GAGCATCACTCCGAGCTTAGCTGGGACCTACCTGAACTACCTGGGCCCCCTTTCAAC 304
QY 300 GAGCAGACTTCAACCTCCCGCTCGCGGCGGAGACTCTGCCAGGGCCACTGTTGAC 359
Db 305 GAGCCTGACTTCAATCTCTCGACTGGGGCGAGAACTCTGCCAGGGCCAGCTGTTGAC 364
QY 360 TTGGAGTGTGGCGAAGCTCAATGACAACTCGCGGCTGACCCAGAACTATGAGGCTAC 419
Db 365 TTGGAGTGTGGCGAAGCTCAATGACAACTCGCGGCTGACCCAGAACTATGAGGCTAC 424
QY 420 AGCCACTTCTGTGTACTGTTGGTGGGACCTATCTGAACTACTGGTGTCTGAAGGAG 479
Db 425 AGTCACTCTCTGTGTACTGTTGGTGGGACCTCAACCGCTCAGGCTGCACTGAACTCCGA 484
QY 480 CGCAGCTGGGCCACTTCTGACAGCTCCAGGCTCGCGGCTGACCCAGAACTACGAGGCTC 539
Db 485 CGTACCTGGGCCACTTCTGACAGCTCCAGGCTCGCGGCTGACCCAGAACTATGAGGCTC 544
QY 540 ATGGAGCTCTGGGCTACCCACTGCGCCAGCGCTCGCTGGGACTGAACTTGGACT 599
```

Db 545 ATGGGAGCGCTTGGCTACCCAGCTGCCAGCCTCTGCCAGGACTGAGCCAGCCTGGGCC 604  
Qy 600 CTGGCCCTGCCACAGTACTCTCCAGAGATGACGACTTCTGGCTGCTGAAGGAG 659  
Db 605 CTGGCCCTGCCACAGTACTCTCCAGAGATGAGTACTCTGGCTGCTGAAGGAG 664  
Qy 660 CTGACAGCTGGCTGGCTGGCCGCAAGGACTTCAACCGGCTCAAGAGAGATGAG 719  
Db 665 CTGACAGCTGGCTGGCTGGCCGCAAGGACTTCAACCGGCTTAAGAGAGATGAG 724  
Qy 720 CTCCAGAGCTGAGTCACTCCCTGCACTGGGGGCTCATGGCTTCTGACTTCTGACTT 778  
Db 725 CTCCAGAGCTTCACTGACTCACTGCACTTGGAGGCAATGGTTCTGACTTCTGACCT 783

## RESULT 7

US-09-016-534-4  
; Sequence 4, Application US/09016534  
; Patent No. 6143874  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; APPLICANT: ELLIOTT, GARY S.  
; APPLICANT: SARMIENTO, ULLA  
; APPLICANT: SENALDI, GIORGIO  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,534  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442B  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 819 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 95..769  
; FEATURE:  
; NAME/KEY: mat peptide  
; LOCATION: 176..769  
; FEATURE:  
; NAME/KEY: sig peptide  
; LOCATION: 95..175  
; US-09-016-534-4

Query Match 84.0%; Score 669.4; DB 3; Length 819;  
Best Local Similarity 92.0%; Pred. No. 7.7e-151;  
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
Qy 1 ATTTAAAGCTTCGCGGAGCCGCGCTCCGCTCCCACTCCGCGGAGGAGGAG 60  
|||||

Db 5 ATTTAAAGCTTCGCGGAGCCGCGCTCCGCTCCCACTCCGCGGAGGAG 64  
Qy 61 CCGCACCCGCGCGGCCAG-CCCCAGCCCCATGACCTCCGAGCAGGAGGAGCTCGTGGGG 119  
Db 65 CCGCGCCGCGCGCGGCCAGCCCCCATGACCTCCGAGCAGGAGGAGCTCGTGGGG 124  
Qy 120 ATGTTAGCGTCTGTGCAAGTCTGCGGCACTCTCCCTGCAAGTGCAGCTCTCAATCGC 179  
Db 125 ATGTTAGCTTGCCTATGCAAGTCTGCGGCACTCTCCCTGCAAGTGCAGCTCTCAATCGC 184  
Qy 180 ACAGGGAGCCAGGGCTGGGCCCTCCATCAGAAAACTATGACTCAACCGCTACCTG 239  
Db 185 ACAGGAGATCAGGGCTGGGCCCTCCATCAGAAAACTATGACTCAACCGCTACCTG 244  
Qy 240 GAGCACCACCTCCGAGCTTGGCTGGGACCTATCTGAACCTACTTGGGCCCCCTTTCAAC 299  
Db 245 GAGCATCAACTCCGAGCTTGGCTGGGACCTACTTGGGCCCCCTTTCAAC 304  
Qy 300 GAGCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCACTGTGTGAC 359  
Db 305 GAGCTGACTTCAATCTCTCGACTGGGGGAGAACTCTGCCAGGGGCACTGTGTGAC 364  
Qy 360 TTGAGGTGTGGGAGAGCTCAATCAAAAATGCGGCTGACCCAGAACTACGAGGCTAC 419  
Db 365 TTGGAAGTGTGGGAGAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGCTAC 424  
Qy 420 AGCCACCTTCTGTGTACTTGGCTGGGCTCAACCGCTCAGGCTGCCACTGTGAGCTGCGC 479  
Db 425 AGTCACTTCTGTGTACTTGGCTGGGCTCAACCGCTCAGGCTGCCACTGTGAGCTGCGC 484  
Qy 480 CGCAGCTTGGGCACTTCTGCAACAGCTTCCAGGGGCTCTGGGAGCAATTTGGGGGCTC 539  
Db 485 CGTAGCTTGGGCACTTCTGTACAGGCTTCCAGGGGCTCTGGGAGCAATTTGGAGGTGTC 544  
Qy 540 ATGGCAGCTTGGGCTTACCACTGCCAGCGCTGCCCTGGGAGTCAACCCACTTTGGACT 599  
Db 545 ATGGCAGCTTGGGCTTACCACTGCCAGCGCTGCCCTGGGAGTCAACCCACTTTGGGCT 604  
Qy 600 CTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 659  
Db 605 CTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 664  
Qy 660 CTGACAGCTGGCTGTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 719  
Db 665 CTGACAGCTGGCTGTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 724  
Qy 720 CTCCAGAGCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTTCTGACCTT 778  
Db 725 CTCCAGAGCTTCACTGACTCACTGCACTTGGAGGCAATGGTTCTGACTTCTGACCTTCTGACCT 783

## RESULT 8

US-08-792-019B-3  
; Sequence 3, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/792,019B

```
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED"
; OTHER INFORMATION: REGION OF >1KB"
US-08-792-019B-3

Query Match 65.7%; Score 523.4; DB 1; Length 5087;
Best Local Similarity 99.8%; Pred. No. 6.9e-116;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 271 ATCTGAACCTACTGGGCCCCCTTTCAACAGCAGCAGACTTCAACCTCCCGCCTGGGG 330
Db 3363 AGCTGAACCTACTGGGCCCCCTTTCAACAGCAGCAGACTTCAACCTCCCGCCTGGGG 3422

QY 331 CAGAGACTCTGCCAGAGGCGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAAAC 390
Db 3423 CAGAGACTCTGCCAGAGGCGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAAAC 3482

QY 391 TCGGCTGACCCAGAACTACAGAGGCGCTACAGCCACTTCTGTGTTACTTGGTGGCTCA 450
Db 3483 TCGGCTGACCCAGAACTACAGAGGCGCTACAGCCACTTCTGTGTTACTTGGTGGCTCA 450

QY 451 ACCGTGAGGCTGCCAGCTGTGAGCTGCGCGCTCATGGAGCTTGGGCTACCCAGCTCC 510
Db 3543 ACCGTGAGGCTGCCAGCTGTGAGCTGCGCGCTCATGGAGCTTGGGCTACCCAGCTCC 3602

QY 511 AGGGCTGCTGGGCGAGCACTGGGCGCTCATGGAGCTTGGGCTACCCAGCTCC 570
Db 3603 AGGGCTGCTGGGCGAGCACTGGGCGCTCATGGAGCTTGGGCTACCCAGCTCC 3662

QY 571 CGCTGCTGGGAGTGAACCCACTTGGAGCTTGGGCTTCCGCGCTTCCCTCCAGA 630
Db 3663 CGCTGCTGGGAGTGAACCCACTTGGAGCTTGGGCTTCCGCGCTTCCCTCCAGA 3722

QY 631 AGATGAGCAGCTTCTGGCTGTGAAGAGAGCTGCAGACCTGGCTGTGGCGCTCGGCAAG 690
Db 3723 AGATGAGCAGCTTCTGGCTGTGAAGAGAGCTGCAGACCTGGCTGTGGCGCTCGGCAAG 3782

QY 691 ACTTCAACCGGCTCAAGAGAGAGTGCAGCTTCCAGAGCTGCAGCTACCCAGCTGG 750
Db 3783 ACTTCAACCGGCTCAAGAGAGAGTGCAGCTTCCAGAGCTGCAGCTACCCAGCTGG 3842

QY 751 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCTGCTCCCC 795
Db 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCTGCTCCCC 3887

RESULT 9
US-08-988-819-3
; Sequence 3, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
```

```
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED"
; OTHER INFORMATION: REGION OF >1KB"
US-08-988-819-3
```

```
Query Match 65.7%; Score 523.4; DB 3; Length 5087;
Best Local Similarity 99.8%; Pred. No. 6.9e-116;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 271 ATCTGAACCTACTGGGCCCCCTTTCAACAGCAGCAGACTTCAACCTCCCGCCTGGGG 330
Db 3363 AGCTGAACCTACTGGGCCCCCTTTCAACAGCAGCAGACTTCAACCTCCCGCCTGGGG 3422

QY 331 CAGAGACTCTGCCAGAGGCGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAAAC 390
Db 3423 CAGAGACTCTGCCAGAGGCGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAAAC 3482

QY 391 TCGGCTGACCCAGAACTACAGAGGCGCTACAGCCACTTCTGTGTTACTTGGTGGCTCA 450
Db 3483 TCGGCTGACCCAGAACTACAGAGGCGCTACAGCCACTTCTGTGTTACTTGGTGGCTCA 3542

QY 451 ACCGTGAGGCTGCCAGCTGTGAGCTGCGCGCTCATGGAGCTTGGGCTACCCAGCTCC 510
Db 3543 ACCGTGAGGCTGCCAGCTGTGAGCTGCGCGCTCATGGAGCTTGGGCTACCCAGCTCC 3602

QY 511 AGGGCTGCTGGGCGAGCACTGGGCGCTCATGGAGCTTGGGCTACCCAGCTCC 570
Db 3603 AGGGCTGCTGGGCGAGCACTGGGCGCTCATGGAGCTTGGGCTACCCAGCTCC 3662

QY 571 CGCTGCTGGGAGTGAACCCACTTGGAGCTTGGGCTTCCGCGCTTCCCTCCAGA 630
Db 3663 CGCTGCTGGGAGTGAACCCACTTGGAGCTTGGGCTTCCGCGCTTCCCTCCAGA 3722

QY 631 AGATGAGCAGCTTCTGGCTGTGAAGAGAGCTGCAGACCTGGCTGTGGCGCTCGGCAAG 690
Db 3723 AGATGAGCAGCTTCTGGCTGTGAAGAGAGCTGCAGACCTGGCTGTGGCGCTCGGCAAG 3782

QY 691 ACTTCAACCGGCTCAAGAGAGAGTGCAGCTTCCAGAGCTGCAGCTACCCAGCTGG 750
Db 3783 ACTTCAACCGGCTCAAGAGAGAGTGCAGCTTCCAGAGCTGCAGCTACCCAGCTGG 3842

QY 751 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCTGCTCCCC 795
Db 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCTGCTCCCC 3887
```

```

RESULT 10
US-09-016-534-3
; Sequence 3, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULLA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,534
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQ
; OTHER INFORMATION: REGION OF >1KB"
; US-09-016-534--3
;

```

	Query Match	65.7%	Score 523.4	DB 3	Length 5087
	Best Local Similarity	99.8%	Pred. No. 6.9e-116		
	Matches 524	Conservative 0	Mismatches 11	Indels 0	Gaps 0
Qy	271	ATCTGAACCTACTCGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGGGG	330		
Db	3363	AGCTGAACCTACTGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGGGG	3422		
Qy	331	CAGAGACTCTGCCACAGGGGCCACTGTTGACTTTGGAGGTTGGCGAAGCCCTCAATGACAAAC	390		
Db	3423	CAGAGACTCTGCCACAGGGGCCACTGTTGACTTTGGAGGTTGGCGAAGCCCTCAATGACAAAC	3482		
Qy	391	TGCGGCTGACCCAGAACTACAGAGGCCCTACAGGCACCTCTGTGTGTACTTTCGTGGCCCTCA	450		
Db	3483	TGCGGCTGACCCAGAACTACAGAGGCCCTACAGGCACCTCTGTGTGTACTTTCGTGGCCCTCA	3542		
Qy	451	ACCGTCAGGCTGCCACTGCTGACTTGGCGCCGACGCTGGCCCACTTCTGCACAGGCTCC	510		
Db	3543	ACCGTCAGGCTGCCACTGCTGACTTGGCGCCGACGCTGGCCCACTTCTGCACAGGCTCC	3602		
Qy	511	AGGGCCCTGCTGGGCGACGATTGCGGGCGTTCATGGCAGCTCTGGGCTACCCACTGCCCCAGC	570		
Db	3603	AGGGCCCTGCTGGGCGACGATTGCGGGCGTTCATGGCAGCTCTGGGCTACCCACTGCCCCAGC	3662		

Qy	571	CGTGCCTGGAGCTGAACCCACTTGGACTCCTGGCCCTGCCACAGTGAATTCCTCCAGA	630
Db	3663	CGCTGCCTGGGACTGAACCCACTTGGACTCCTGGCCCTGCCACAGTGAATTCCTCCAGA	3722
Qy	631	AGATGGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGG	690
Db	3723	AGATGGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGG	3782
Qy	691	ACTTCAACCGGCTCAAGAAAGAGATGCAGCCTCCAGCAGCTGCAGTCAACCTGCACCTGG	750
Db	3783	ACTTCAACCGGCTCAAGAAAGAGATGCAGCCTCCAGCAGCTGCAGTCAACCTGCACCTGG	3842
Qy	751	GGGCTCATGGCTTCTGACTTCTGACCTTCTCCTCTTGGCTCCGCC	795
Db	3843	GGGCTCATGGCTTCTGACTTCTGACCTTCTCCTCTTGGCTCCGCC	3887

```

RESULT 11
US-09-106-182-7
; Sequence 7, Application US/09106182
; Patent No. 6046035
; GENERAL INFORMATION:
; APPLICANT: Shi, Yangu
; APPLICANT: Ruben, Steve
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,182
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/051,053
; FILING DATE: 30-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF385
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-106-182-7

```

Query Match	10.1%	Score 80.2;	DB 3;	Length 396;
Best Local Similarity	96.5%	Pred. No. 1.4e-10;		
Matches a2: Conservative	0;	Mismatches 3;	Indels 0;	Caps 0;

QY	711	AAGATGACAGCCTTCAGCAGCTGAGTCAACCTGCACCTGGGGGCTCATCGCTTCTGACTT	770
DB	4	ACGAGGAGAGCCTCAGCAGCTGAGTCAACCTGCACCTGGGGGCTCATCGCTTCTAACTT	63
QY	771	CTGACCTTCTCCTCTTCGCTCCCC	795
DB	64	CTGACCTTCTCCTCTTCGCTCCCC	88

RESULT 15  
US-09-568-480-1  
; Sequence 1, Application US/09568480  
; Patent No. 6355458  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross

RESULT 15  
US-09-568-480-1  
; Sequence 1, Application US/09568480  
; Patent No. 6355458  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Liigon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross

us-09-931-704-1.rni

Sun Feb 2 08:31:38 2003

```
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match 5.6%; Score 44.8; DB 4; Length 68750;
Best Local Similarity 50.0%; Pred. No. 0.12; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 112;

QY 456 CAGGCTGCCACTGCTGAGCTGCGCGCGAGCTGGCCCACTTTCGACCCAGCTCCAGGGC 515
Db 57992 CAGGTAGCGAGTCCGGAGCCGGGGGACGTCGCGGCCCTTCGGATCCACGCCCGGGC 58051

QY 516 CTGCTGGGCGACATGCGGGCGTCATGGCAGCTCTGGGCTACCCACTGCCCGCCCGCTG 575
Db 58052 GTGCTGCCCGGATCGGGCGCGTCGAGACCCCGGGGAGGTGAACTCGCCGCGCTGCGC 58111

QY 576 CTGGGACTGAACCACTTGGACTCTCTGGCCCTGCCACAGTGACTTCTCCAGAAGATG 635
Db 58112 GCCCGGCTTCATGCCCGCGTCCCGCTCGCGCTATCTATGTGGCTCGCCGAGATGGG 58171

QY 636 GACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCG 679
Db 58172 CTTCAATACGCCCGCGCTTGCAGGGGCTCGCGAGCTGTGGCG 58215
```

Search completed: February 1, 2003, 08:46:09  
Job time : 240.683 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:22 ; Search time 14.5 Seconds  
(without alignments)  
1491.741 Million cell updates/sec

Title: US-09-931-704-2

Perfect score: 1226

Sequence: 1 MDLRAGDSWGLACLCTVLM.....KKKMQPPAAAVTLHLGAHGF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	124.5	10.2	203	2 JC4645	cardiotrophin-1 -
2	118.5	9.7	203	2 I49153	cardiotrophin-1 -
3	92	7.5	332	2 G86182	hypothetical prote
4	91.5	7.5	201	2 G02312	cardiotrophin-1 -
5	91.5	7.5	1182	2 I48378	hairless protein -
6	91	7.4	195	2 JH0680	ciliary neurotroph
7	89.5	7.3	1313	2 T38943	probable guanine n
8	89	7.3	733	2 D83588	conserved hypotet
9	87	7.1	955	2 T00247	zinc finger protei
10	87	7.1	1561	2 T00248	zinc finger protei
11	86.5	7.1	640	2 T34916	transferase - stre
12	85.5	7.0	560	2 C38604	poly(3-hydroxyalka
13	84.5	6.9	389	2 F72511	probable cystathio
14	84	6.9	200	2 AD3633	hypothetical prote
15	84	6.9	411	2 F75439	probable cell wall
16	84	6.9	559	2 G71327	probable apolipop
17	83	6.8	452	2 A35602	tailless (tll) pro
18	82.5	6.7	542	2 A82965	hypothetical prote
19	82.5	6.7	723	2 B38749	3-phosphatidylinos
20	82	6.7	1179	2 T04584	TMW resistance pro
21	81	6.6	479	2 C97508	blue-light photore
22	81	6.6	479	2 A42726	DNA photolase [im
23	80.5	6.6	353	2 AB1823	hypothetical prote
24	80.5	6.6	512	2 S21171	activin receptor S
25	80	6.5	200	1 UNRTCF	ciliary neurotroph
26	80	6.5	460	2 D75493	cell division cycl
27	80	6.5	2261	1 A42548	genome polyprotein
28	79	6.4	812	2 T34180	hypothetical prote
29	79	6.4	1220	2 AD0125	exodeoxyribonuclea

protein-tyrosine k  
hypothetical prote  
conserved hypotet  
yes-associated pro  
hypothetical prote  
CDA peptide synthe  
hypothetical prote  
hypothetical prote  
extracellular solu  
65k early nonstruc  
65k early nonstruc  
hypothetical prote  
conserved hypotet  
hypothetical prote  
conserved hypotet

#### ALIGNMENTS

##### RESULT 1

JC4645

cardiotrophin-1 - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Jun-2000

C/Accession: JC4645

R/Ishikawa, M.; Saito, Y.; Miyamoto, Y.; Kuwahara, K.; Ogawa, E.; Nakagawa, O.; Harada

Biochem. Biophys. Res. Commun. 219, 377-381, 1996

A/Title: cDNA cloning of rat cardiotrophin-1 (CT-1): Augmented expression of CT-1 gene

A/Reference number: JC4645; MUID:96193659; PMID:8604995

A/Accession: JC4645

A/Molecule type: mRNA

A/Residues: 1-203 <ISH>

A/Cross-references: DDBJ:D78591; NID:91256926; PIDN:BAA11427.1; PID:91256927

C/Genetics:

A/Gene: CT-1

C/Keywords: cardiac muscle; cytokine; heart

Query Match 10.2%; Score 124.5; DB 2; Length 203;

Best Local Similarity 28.1%; Pred. No. 0.00045;

Matches 50; Conservative 24; Mismatches 85; Indels 19; Gaps 5;

QY 40 IQKTYDLTRVLEHQRLSLAGTYLNYLGPFPNEPDPFNPRL---GAETLPRAVTDLVWRS 96

Db 27 IRQTHNLARLLTKYADQLLEVEYVQQQEPFGLPGFSPRLPLAGLSGPAPSHAGLPV--- 83

QY 97 LNDKLRITQNYEAYSHLLCYLRGLNRQAA-----TAE LRSLAHFTCSLQGLLSIAGVM 151

Db 84 ---SERLRQDAALSAIPALIDAVRRRQALNPAPRLRLSRLEDAARQVRAAGAAVTVL 140

QY 152 AALGY----PLPQLPGTEPTWTGPAHSDFLOKMDDFWLLKELQTLWLRSAKDPNRL 205

Db 141 AALGAAARGVPPEPV-ATSAFTSNSAAGVFSKVLGLHVCGLYGEWVSRTEGDLGQL 197

##### RESULT 2

I49153

cardiotrophin-1 - mouse

C/Species: Mus musculus (house mouse)

C/Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999

C/Accession: I49153

R/Pennica, D.; King, K.L.; Shaw, K.J.; Luis, E.; Rullamas, J.; Luoh, S.; Darbonne, W.C.

Proc. Natl. Acad. Sci. U.S.A. 92, 1142-1146, 1995

A/Title: Expression cloning of cardiotrophin 1, a cytokine that induces cardiac myocyte

A/Reference number: I49153; MUID:95166785; PMID:7862649

A/Accession: I49153

A/Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ

A/Molecule type: mRNA

A/Residues: 1-203 <RES>

A/Cross-references: EMBL:U18366; NID:g710331; PIDN:AAC52173.1; PID:g710332

C/Genetics:

A/Gene: ctfl

C;Keywords: cytokine; glycoprotein  
F;164/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.7%; Score 118.5; DB 2; Length 203;  
Best Local Similarity 28.8%; Pred. No. 0.0016;  
Matches 51; Conservative 21; Mismatches 88; Indels 17; Gaps 5;

QY 40 IQKTYDTRYLEHQLRSLAGTYLNYLGGPFNEPDPNPRRL---GAETLPRAVDLEWRS 96  
DB 27 IROTHNLARLLTKYAEQLLEEVQOGEFGLGFCFPRLPLAGLSGAPSHAGLPV-- 83  
QY 97 LNDKRLTONYAYSHLLCYLRLGNROAA-----TAE LRSLAHFCTSLQGLLSGIAGVM 151  
DB 84 ---SERLRQDAALSVLPALLDAVRRRQAEINPRAPRLRLSLDAARQVRAALGAAVETVL 140  
QY 152 AALGYPL--POPLGTEPT-WTPGPAHSDFLOKMDDFLLKELQTLWLRSAKDFNRL 205  
DB 141 AALGAAARGPPEPTVATLFTANSTAGIFS AKVLGFHVCGLYGEWVSRTGDLGQL 197

## RESULT 3

G86182  
hypothetical protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: G86182  
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huijar, L.  
Nature 408, 816-820, 2000  
C;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: G86182  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-332 <STO>  
A;Cross-references: GB:AE005172; NID:g7211973; PIDN:AAF40444.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1

Query Match 7.5%; Score 92; DB 2; Length 332;  
Best Local Similarity 26.4%; Pred. No. 0.84;  
Matches 58; Conservative 24; Mismatches 72; Indels 66; Gaps 13;

QY 23 PAVPALNRTGDPGPGPSIQKTYDTRYLEHQLRSLAGTYLNY-----PGP----- 64  
DB 113 PSVTAGNLGYP-PRSF--TYDPGYEQRMESLLQOQFIRENPQIRPLRGLGSPVG 169  
QY 65 LGPPNPNPNPRLGAETLPRAVDLEWRSNDKRLRLTONYAYSHLLCYLRLGNROA 124  
DB 170 LGPIRASQFLQPRVAPP--PTSILD-----TSNRKAR-----SKDGLAVVRG--RKV 215  
QY 125 ATAE LRSL-----AHFCTSLQGLLSGIAGVMALGYPLPPLP---GTEPTWT 170  
DB 216 RITEGSSSYSLGRSLKNGAHV-----GIQQRSGIMK-----PLPKPLVDLTTSVTP 266  
QY 171 PGPAHSDFLOKMDDFLLKELQTLWLRSAKDFNRLKKMKQ 210  
DB 267 DDPDESAEDKDEEAVKQL-----SEKDL--LKEHIE 298

## RESULT 4

G02312  
cardiotrophin-1 - human  
C;Species: Homo sapiens (man)  
C;Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 17-Jul-1998  
C;Accession: G02312  
R;Wood, W.I.;

submitted to the EMBL Data Library, December 1995  
A;Reference number: H01035

A;Accession: G02312  
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA  
A;Residues: 1-201 <WOO>  
A;Cross-references: EMBL:U43030; NID:g1151149; PID:g1151150

C;Genetics:  
A;Gene: GDB:CTF1; CT-1  
A;Cross-references: GDB:567078  
A;Map position: 1p22-1p22

Query Match 7.5%; Score 91.5; DB 2; Length 201;  
Best Local Similarity 25.9%; Pred. No. 0.49;  
Matches 45; Conservative 24; Mismatches 92; Indels 13; Gaps 5;

QY 40 IQKTYDTRYLEHQLRSLAGTYLNYLGGPFNEPDPNPRRLGAETLPRAVDLEWRSND 99  
DB 27 IROTHSLAHLTKYAEQLLEEVQOGEFGLGFCFPRLPVAGL-SAPAPSHAGLPVHE 85  
QY 100 KRLRTQNYEAYSHLLCYLRLGNROAA-----TAE LRSLAHFCTSLQGLLSGIAGVMAAL 154  
DB 86 RLRL--DAALALPPLLDVAVCRQAEINPRAPRLRLRLDAARQARALGAAVEALLAAL 143  
QY 155 GYPLPQPLPGTEP---TWTPGPAHSDFLOKMDDFLLKELQTLWLRSAKDFNRL 205  
DB 144 G--AANRGPRAPPAATASAAATGVPKAVGLRLVCGLYREWLRSRTGDLGQL 195

## RESULT 5

I48378  
hairless protein - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C;Accession: I48378  
R;Cachon-Gonzalez, M.B.; Penner, S.; Coffin, J.M.; Moran, C.; Stoye, J.P.  
Proc. Natl. Acad. Sci. U.S.A. 91, 7717-7721, 1994  
A;Title: Structure and expression of the hairless gene of mice.  
A;Reference number: I48378; MUID:94329587; PMID:8052649  
A;Accession: I48378  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1182 <RES>  
A;Cross-references: EMBL:Z32675; NID:g531706; PIDN:CAA83587.1; PID:g531707

Query Match 7.5%; Score 91.5; DB 2; Length 1182;  
Best Local Similarity 23.7%; Pred. No. 4.6;  
Matches 70; Conservative 25; Mismatches 99; Indels 101; Gaps 17;

QY 3 LRAGDSGMLACLCTVLVHLPAVPALNRTGDPG-----PGP----- 38  
DB 813 LRAGS--GLRKGL-----SLPLSPVRLTSLPGALLWLQEPKPHGLFQEHWRGQPV 865  
QY 39 ---SIQKTYDTRYLEHQLRSLAGTY--LNVIGPP-----FNEPDPNPRLG 80  
DB 866 LVSGIQKTLRLSLGMEALGTGGQVQTLTALGPPQPTNLDSTAFWEGSFHPETRP---- 921  
QY 81 AETLPRAVDLEWRSNDK-LRLTON-----YEAYSHLLCYL-RGLNRQAAT 126  
DB 922 ---KLDEGSV-LLHRTLTGDKDASRVONLASSLPDPEYCAHQKLNLAASYLPGLTLHLE 978  
QY 127 AEL-----RRSLAHFCTSLQGLLSGIAGVMALGYPLPQPLGTEPTWTGPAH 175  
DB 979 PQLWAAYGVNSHRGHGTLKNCVEVSDLSILVHAEAL-----PPWY--RAQ 1024  
QY 176 SDFLOKMD--DFWLLKELQTLWLR--SAKDFNRLKKMKQ--PAAAVTLHLGHG 224  
DB 1025 KDFLSGLDGEGLWSPGSGTSTVWHVFRQAQRIIRFLQMVCPAGATLEPGAPG 1079

## RESULT 6

JH0680  
ciliary neurotrophic factor - chicken



N;Alternate names: growth-promoting activity protein  
C;Species: Gallus gallus (chicken)  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jul-2000  
C;Accession: J06080; PQ0057

R;Leung, D.W.; Parent, A.S.; Cachianes, G.; Esch, F.; Coulombe, J.N.; Nikolic, K.; Eickbush, S.H.; 1045-1053, 1992  
A;Title: Cloning, expression during development, and evidence for release of a trophic factor from embryonic peripheral neurons  
A;Reference number: J06080; PMID:92304573; PMID:1610564  
A;Accession: J06080  
A;Molecule type: mRNA  
A;Residues: 1-195 <LEU>  
A;Cross-references: GB:M80827; NID:g211822; PIDN:AAA48784.1; PID:g211823  
A;Experimental source: eye  
R;Eckstein, F.P.; Esch, F.; Holbert, T.; Blacher, R.W.; Nishi, R.  
Neuron 4, 623-631, 1990

A;Title: Purification and characterization of a trophic factor for embryonic peripheral neurons  
A;Reference number: PQ0057; PMID:90211978; PMID:2322465  
A;Accession: PQ0057  
A;Molecule type: protein  
A;Residues: 155-166, 'X', 168-175 <ECK>  
A;Experimental source: sciatic nerves  
C;Comment: This is a neurotrophic protein.  
C;Superfamily: ciliary neurotrophic factor  
C;Keywords: growth factor

Query Match 7.4%; Score 91; DB 2; Length 195;  
Best Local Similarity 27.3%; Pred. No. 0.53;  
Matches 51; Conservative 21; Mismatches 83; Indels 32; Gaps 9;

QY 46 LTRYLEHQLRSLAGTYLNYLGGPPNEPDPFPPRIGATLPRATVDLEVRSLNDKRLQ 105

DB 23 LARKMRSDVTDLLDIYVERQG-----LDASISVAADVGVPTAAV--ERWAEQTGTQRLD 75

QY 106 N---YEAYSHLLCYLRLNQA---ATAELRSLA-----HPTSLQGLGSIAGVMA 152

DB 76 NLAAAYRFTLLAQMEEQELGDTDAELGPALAAMLLQVSFVYHLELL-----ELE 130

QY 153 ALGYPLPQPLPGTEPTTPGPAH--SDFLQKMDDFWLLKELQTLWRSAKDFNRLKKMQP 211

DB 131 SRGAPAE---GSEP---PAPPRLSLFQKRLGLRLVRLAQWAVRSVRDLRLQSKHGP 184

QY 212 PAAAVTL 218

DB 185 SGAALGL 191

## RESULT 7

T38943  
probable guanine nucleotide binding protein - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C;Accession: T38943

R;Skelton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, May 1997  
A;Reference number: Z21819  
A;Accession: T38943

A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1313 <SKE>  
A;Cross-references: EMBL:Z95396; PIDN:CA808769.1; GSPDB:GN00066; SPDB:SPAC57A7.11  
A;Experimental source: strain 972h-; cosmid c57A7  
C;Genetics:  
A;Gene: SPDB:SPAC57A7.11  
A;Map position: 1

Query Match 7.3%; Score 89.5; DB 2; Length 1313;  
Best Local Similarity 22.8%; Pred. No. 8;  
Matches 61; Conservative 32; Mismatches 69; Indels 105; Gaps 14;

QY 15 LCTVLMHLPVAV--PALNR-----TGDPGPGPSI-----QKTYDLTRYLEHQLRS 56

DB 811 LAFLLQHLPALHAKSLSKDTNTSVTSDPKPHFPVPSVSENKILNRSFSLTRSLKGLALS 870

QY 57 LAG-----TYLNYL-----GPP-FNEPDPFNPPRLGAETLPRATVDL 91  
DB 871 LAGSDRASELLSNGENKPAESNLHLTSKAVPGPAPFAFNELEY-----QSELD 919

QY 92 EVWRSND-----KLRLTONYEAYSHLLCYLRLG---NRQATAELRSLAHECTS 139

DB 920 PLTSYLPDWSRKYFTFQMPRPNEDEPGS--ICYNQRLWRNRNEKLIYTRPLAESTN 977

QY 140 -----LQGLLSIAGVMAALGYPLPQPLPGTEPTTPGPAHSDFLQKMDDFWLLKE--- 190

DB 978 GRNNQMLTFNTTIA-----PRKLMFHFQEDQLITLGDKDI 1013

QY 191 LQTLWRSAKDFNRLKKMQPPAAAVT 217

DB 1014 IQVWDRR-----NRCLNSFKTSASATT 1036

## RESULT 8

DB3588

conserved hypothetical protein PA0454 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: D83588

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Li

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

A;Reference number: A82950; PMID:20437337; PMID:10984043

A;Accession: D83588

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-733 <STO>

A;Cross-references: GB:AE004483; GB:AE004091; NID:g9946313; PIDN:AAG03843.1; GSPDB:GN0

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA0454

C;Superfamily: hypothetical protein HI1680

Query Match 7.3%; Score 89; DB 2; Length 733;  
Best Local Similarity 26.8%; Pred. No. 4.3;  
Matches 58; Conservative 26; Mismatches 66; Indels 68; Gaps 13;

QY 28 LNRTGDPGPGPSIQKTYDLTRYLE-----HOLRSLAGTYLNYLGGPPNEPD--FNPP 77

DB 230 LNRLGHGRPGKV-----SRYLKLYPIAQDVHERASSSHYPYNNRLAEAFHSDVLRQC 283

QY 78 RL-----GAETLPRATVDLEVRSLNDKRLTLQNYEAYSHLLCYLRLGRLNQAATAELRR 131

DB 284 RLNNQOGKACQALARA-----IRLRQPPD-YA-----DRELALEDLQA 320

QY 132 SLAHF-----CTSLQGLLSIAGVMAALGYPLPQPLPGTEPTTPGPAHSDFLQKMDDFWL 187

DB 321 SLEHLRQOSNPAPKGLLRSL-GALANLTLDRKLAGA-----SNPDIADEQDSAL 371

QY 188 LKELQTLWRSAKD-FNRLKKMQPPAAAVTLHLGAGH 224

DB 372 LDRSP-----RSLKDAFERLRQQLTP-----TSLLFRHG 400

## RESULT 9

T00247

zinc finger protein wiz - mouse

C;Species: Mus musculus (house mouse)

C;Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 05-Nov-1999

C;Accession: T00247

R;Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiosaka, S.; Wanaka, A.; Taniyama, M.

submitted to the EMBL Data Library, March 1998

A;Description: Molecular cloning and distinct developmental expression pattern of spli

A;Reference number: Z14130

A;Accession: T00247

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-955 <MAT>  
A;Cross-references: EMBL:AB012266; NID:d1227741; PIDN:BAA32791.1; PID:d1033757  
A;Experimental source: brain  
C;Genetics: wiz  
A;Gene: wiz

Query Match 7.1%; Score 87; DB 2; Length 955;  
Best Local Similarity 27.8%; Pred. No. 9.2;  
Matches 35; Conservative 19; Mismatches 48; Indels 24; Gaps 7;

QY 22 LPAPALNRTGPGGPGPSIQKTYDLYLRYLHQLRSLAG---TYLVNIGP-----PFNEPD 73

Db 615 LPLSPLASRPGKPGAGT-----QVPR--ELSLSPITGSKPSAASVILGVPVATKRPQEDR 667

QY 74 FNPRLGAETLPRATVDLEWRSINDKRLRTQNYEAYSHLLCYLRGL---NROAATAELR 130

Db 668 FLPAEVKAKTYIQTELPPFKA-KTLHEK-----TSHSSTEACCELCGLYFENRKALASHAR 721

QY 131 RSLAHF 136

Db 722 AHLRQF 727

#### RESULT 10

T00248

zinc finger protein wizL - mouse

C;Species: Mus musculus (house mouse)

C;Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 05-Nov-1999

C;Accession: T00248

R;Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiozaka, S.; Wanaka, A.; Tohyama, M.

submitted to the EMBL Data Library, March 1998

A;Description: Molecular cloning and distinct developmental expression pattern of splice

A;Reference number: Z14130

A;Accession: T00248

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1561 <MAT>

A;Cross-references: EMBL:AB012265; NID:d1227740; PIDN:BAA32790.1; PID:d1033756

A;Experimental source: brain

C;Genetics: wiz

A;Gene: wiz

Query Match 7.1%; Score 87; DB 2; Length 1561;  
Best Local Similarity 27.8%; Pred. No. 17;  
Matches 35; Conservative 19; Mismatches 48; Indels 24; Gaps 7;

QY 22 LPAPALNRTGPGGPGPSIQKTYDLYLRYLHQLRSLAG---TYLVNIGP-----PFNEPD 73

Db 1221 LPLSPLASRPGKPGAGT-----QVPR--ELSLSPITGSKPSAASVILGVPVATKRPQEDR 1273

QY 74 FNPRLGAETLPRATVDLEWRSINDKRLRTQNYEAYSHLLCYLRGL---NROAATAELR 130

Db 1274 FLPAEVKAKTYIQTELPPFKA-KTLHEK-----TSHSSTEACCELCGLYFENRKALASHAR 1327

QY 131 RSLAHF 136

Db 1328 AHLRQF 1333

#### RESULT 11

T34916

transferase - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 02-Sep-2000

C;Accession: T34916

R;Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, January 1998

A;Reference number: Z21558

A;Accession: T34916

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-640 <OLI>

A;Cross-references: EMBL:AL021409; PIDN:CAA16181.1; GSPDB:GN00070; SCODEB:SC3F7.10

A;Experimental source: strain A3 (2)

C;Genetics:

A;Gene: SCODEB:SC3F7.10

C;Superfamily: glycine C-acetyltransferase homology

F;287-624/Domain: glycine C-acetyltransferase homology <GCA>

Query Match 7.1%; Score 86.5; DB 2; Length 640;  
Best Local Similarity 27.2%; Pred. No. 6.2;  
Matches 50; Conservative 13; Mismatches 64; Indels 57; Gaps 9;

QY 22 LPAPALNRTGPGGPGPSIQKTYDLYLRYLHQL---RSLAGTYLVNIGLPGFFNEPDENP 78

Db 100 LPAPVA-EGTGDPTVEAVAAAMRQTQYRHLGLDADLEG-----E 140

QY 79 LGABTLPRATVDLEWRSINDKRLRTQNYEAYSHLLCYLRGLNRQAATAELRRSLAHFCT 138

Db 141 LGVDSVVTTSVABA---TERLGLT-----GAAPDAAGATTIRALA---D 179

QY 139 SLOGLGSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLOKMDDFWLLKELQTLWLR 198

Db 180 ALRGL-----VAAAAPGTVAPEAAPATGAA-APAPGRSGNAP-----APGADGWDHRS 225

QY 199 AKDF 202

Db 226 MKDF 229

#### RESULT 12

C38604

poly(3-hydroxyalkanoate) polymerase (EC 2.7.7.-) 3 - Pseudomonas oleovorans

C;Species: Pseudomonas oleovorans

C;Date: 23-Aug-1991 #sequence\_revision 23-Aug-1991 #text\_change 08-Sep-2000

C;Accession: C38604

R;Huisman, G.W.; Wink, E.; Meima, R.; Kazemier, B.; Terpstra, P.; Witholt, B.

J. Biol. Chem. 266, 2191-2198, 1991

A;Title: Metabolism of poly(3-hydroxyalkanoates) (PHAs) by Pseudomonas oleovorans.

A;Reference number: A38604; MUID:91115830; PMID:1989978

A;Accession: C38604

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-560 <HUI>

A;Cross-references: GB:M58445; NID:g151441; PIDN:AAA25934.1; PID:g151444

C;Superfamily: poly(3-hydroxyalkanoic acid) synthase phbC

C;Keywords: nucleotidyltransferase

Query Match 7.0%; Score 85.5; DB 2; Length 560;  
Best Local Similarity 30.2%; Pred. No. 6.4;  
Matches 38; Conservative 17; Mismatches 42; Indels 29; Gaps 7;

QY 77 PLGAETLPRATVDLEWRSINDKRLRTQNYEAYSHLLCYLRGLNRQAATAELRRSL--A 134

Db 5 PAKGTPLPATSMNVQ-----NAILGLRGR-----DLISLRNVSQRS---LRHPLHTA 50

QY 135 HFCTSLQGLGSIAGVMAALGYPLPQPLPG-----TEPTWTPGPAHSDFLOKMDDFWLLKE 190

Db 51 HHLALGGLGRV-----ILGDTPLQPNRPDRPFSPTWSQNPFFYRRGLQA----YLA 101

QY 191 LOTWLW 196

Db 102 KQTRLW 107

#### RESULT 13

F72511

probable cystathionine gamma-synthase APE2068 - Aeropyrum pernix (strain K1)

C;Species: Aeropyrum pernix

C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C;Accession: F72511

R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A;Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: F72511  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-389 <RAW>  
A:Cross-references: DDBJ:AF000063; NID:g5105654; PIDN:BAA81078.1; PID:g5105766  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE2068  
C:Superfamily: O-succinylhomoserine (thiol)-lyase

Query Match 6.9%; Score 84.5; DB 2; Length 389;  
Best Local Similarity 22.9%; Pred. No. 5;  
Matches 47; Conservative 19; Mismatches 78; Indels 61; Gaps 8;  
QY 43 TYDLTRYLEHQLRSLAGTYLNYLGGPPNEPDPFPPRLGAETLPRATVDLEWRSLSL----- 97  
Db 107 TYGSTSLLEMLSSITGIEVRLAGPPWED-----LIDLVCWADLIIVES 150  
QY 98 --NDKRLRTQNYEAYSHLLCYLRGLNRQAATAELR-----RSL---AHFCTS 139  
Db 151 MANPTLRVPP-----LSGIYEAGSCGVRVVVDNTFATPIAYRPLERGAHY--S 197  
QY 140 LQGLLSIAGVMAALGYPPLPOPL-PGTEPTWTCGPAHSDFLQKMDDFWLLKELQTLWLWRS 198  
Db 198 LESLTKYIAGHNDVVGSLSGRVEEDLEPLNMWKILGTIMQPIDAY-----LAWRG 249  
QY 199 AKDFNRLKKQWQPPAAATVHLGAH 223  
Db 250 MKTLKARFEAQSAARVAEAWLESH 274

## RESULT 14

AD3633  
hypothetical protein BMEII0989 [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: AD3633  
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, M.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AD3633  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-200 <KUR>  
A:Cross-references: GB:AE008918; PIDN:AAL54231.1; PID:g17985203; GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEII0989  
A:Map position: II

Query Match 6.9%; Score 84; DB 2; Length 200;  
Best Local Similarity 24.3%; Pred. No. 2.4;  
Matches 56; Conservative 22; Mismatches 84; Indels 68; Gaps 13;  
QY 10 GMLACLCTVWLHLPALNRTGDPGPGPSIQKTYDTRYLEHQLRSLAGTYLNLGLGPPF 69  
Db 2 GLLGAAALV---LPSLFAEAKTQQAAPPNATSPHQADVYL---LRGFADIFSTGI----- 51  
QY 70 NEPDFNPRRLGAETLPRATVDLEY-----WRSNDKRLRTQNYEAYSHLLCYLRG--LNR 122  
Db 52 -----DEIGAE-LQAAGVNAHVQGHAAWELVLN--RIVADQKNGHLPVVLIGHSLGA 101  
QY 123 QAA---TAEILRSIAHFTCSLQGLGSIAGVMAALGYPPLPOPLPGT-----EPTW 169  
Db 102 NAAIYIAEELER-----RGIAVDYMATFAATG---PDPLFCNVRRVNVFYFKQHW 149  
QY 170 ----TPGPAHSDFLQKMDDFWLLKELQTLWLWRSKDFNRLKKQWQPPAAA 215  
Db 150 GLPLVPGPRFHGHLENRD-----FSNAKDVGHFNIKQRPLOA 187

## RESULT 15

F75439  
probable cell wall synthesis protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: F75439  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A/5250; MUID:20036896; PMID:10567266  
A:Accession: F75439  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-411 <WHI>  
A:Cross-references: GB:AE001958; GB:AE000513; NID:g6458805; PIDN:AAF10649.1; PID:g6458805  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR1076  
A:Map position: 1

Query Match 6.9%; Score 84; DB 2; Length 411;  
Best Local Similarity 24.6%; Pred. No. 6;  
Matches 56; Conservative 25; Mismatches 77; Indels 70; Gaps 11;  
QY 2 DLRAAGDSWGLACLCTVWLHLPALNRTGDPGPGPSIQKTYDTRYLEHQLRSLAGTY 61  
Db 212 EVLAADRWAIRAELEFSEMSLRPGVPLLLSGG-GRGHYAAAADVLT-----ELGNL-GRA 264  
QY 62 LNYLGPPFPNEPDPFPPRLGAETLPRATVDLEWRSNDKRLRTQNYEAYSHLLCYLRGLN 121  
Db 265 VQVLVPAASRQGE-----GTETIGGATV-----HHLGFRRLDLP 296  
QY 122 RQATAEILRSIAHFTCSLQGLGSIAGV-----MAALGYPLP-QPLPGTEPTWTPGPAH 175  
Db 297 RLAAASDL-----VVVGKAGGLTVAEATALGVPLVIYAPIGQE-----EHN 337  
QY 176 SDFLQKMDDFWLLKELQTLWLW-RSAKDFNRLKKQWQPPAAAATVHLGA 222  
Db 338 ADFLERHG-----AGLWARAHVDVRPLVLRALDPAEHARLSAGA 376

Search completed: January 27, 2003, 15:42:34  
Job time : 16.5 secs







CC similarity).

CC -!- TISSUE SPECIFICITY: Expressed in brain, testis, kidney, stomach,

CC small intestine, liver, and lung. Not detected in heart, skeletal

CC muscle, and spleen.

CC -!- MISCELLANEOUS: Mice homozygous for a null mutation of the CST gene

CC display hindlimb weakness from week 6 of age and subsequently show

CC a pronounced tremor and progressive ataxia. Myelin vacuolation is

CC observed in the cerebellar white matter, diencephalon, brainstem

CC and spinal anterior column. Male mice were infertile due to a

CC blocked spermatogenesis.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AB032940; BAA93009.1; -

DR EMBL; AB032939; BAA93008.1; -

DR EMBL; AK007645; BAB25160.1; -

DR EMBL; BC026806; AAB26806.1; -

DR MGI; MGI:1858277; Gcst

KW Transferase; Transmembrane; Glycoprotein.

FT DOMAIN 1 12

FT TRANSMEM 13 35

FT 1 12

FT 36 423

FT CARBOHYD 66 66

FT CARBOHYD 312 312

FT CARBOHYD 18 18

FT CONFLICT 263 263

FT CONFLICT 271 271

FT CONFLICT 358 358

FT CONFLICT 392 392

FT CONFLICT 398 398

FT CONFLICT 423 423

SQ SEQUENCE 423 AA; 48968 MW; FD54AIA71F4AEE46 CRC64;

Query Match 7.7%; Score 94; DB 1; Length 423;

Best Local Similarity 22.7%; Pred. No. 0.25;

Matches 55; Conservative 23; Mismatches 60; Indels 104; Gaps 12;

QY 44 YDLTRYLEHQRLSLAGTVINY---LGPP-----FNEP----- 72

Db 204 YDPSYNAHYLNLLFFDLGYDSSLDSPRVQEHILEVERFHLVILQEVFDESILVLR 263

QY 73 -----DFNPRLGAETLPRAT-----VDLEVWRSINDKRLTQNYEA 109

Db 264 ELLCWDLEDVLYFKLNARDSPVPLSGELYRRATAWNLLDVLRYRHFN--ASFWRKYEA 321

QY 110 YSHLLCYLRGLNROA-ATAELRS---LAHFCTSLQGLLGSTIA-----GVMAA 153

Db 322 F-----GREMAREVAELRQAEHMRHCIDGGQVAGAEIQDSAMQWPQPLGIKSI 373

QY 154 LGYPLPQLPCTPTWTGPAHSDFLOKMDFFWLKELQ-----TWLRSAXDF 202

Db 374 LGYNLKKSI-----GPQHEQLCRM-----LTPETQYLSDLGALNWLTKLWFLRDF 420

QY 203 NR 204

Db 421 LR 422

RESULT 4

NXFL HUMAN STANDARD; PRT; 619 AA.

AC Q9UBI9; Q9UBI2; Q99799;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Nuclear RNA export factor 1 (Tip associating protein) (Tip-associated protein) (mRNA export factor TAP).

GN NXF1 OR TAP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]\_

RP SEQUENCE FROM N.A., AND MUTAGENESIS.

RC TISSUE=Cervical carcinoma;

RX MEDLINE=99219873; PubMed=1020158;

RA Braun I.C., Rohrbach E., Schmitt C., Izaurralde E.;

RT "TAP binds to the constitutive transport element (CTE) through a novel

RT RNA-binding motif that is sufficient to promote CTE-dependent RNA

RT export from the nucleus.";

RL EMBO J. 18:1953-1965(1999).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99257272; PubMed=10323864;

RA Kang Y., Cullen B.R.;

RT "The human Tap protein is a nuclear mRNA export factor that contains

RT novel RNA-binding and nucleocytoplasmic transport sequences.";

RL Genes Dev. 13:1126-1139(1999).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=99384298; PubMed=10454577;

RA Bear J., Tan W., Zolotukhin A.S., Taberner C., Hudson E.A.,

RA Felber B.K.;

RT "Identification of novel import and export signals of human TAP, the

RT protein that binds to the constitutive transport element of the type

RT D retrovirus mRNAs.";

RL Mol. Cell. Biol. 19:6306-6317(1999).

RN [4]

RP SEQUENCE FROM N.A.

RX TISSUE=Placenta;

RA Strausberg R.;

RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RX TISSUE=Lung;

RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,

RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,

RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,

RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,

RA Isogai T., Sugano S.;

RT "NEDO human cDNA sequencing project ";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE OF 61-619 FROM N.A.

RC TISSUE=Lymphocytes;

RX MEDLINE=97318898; PubMed=9175835;

RA Yoon D.-W., Lee H., Seol W., DeMaria M., Rosenzweig M., Jung J.U.;

RT "Tap: a novel cellular protein that interacts with tip of herpesvirus

RT saimiri and induces lymphocyte aggregation.";

RL Immunity 6:571-582(1997).

RN [7]

RP FUNCTION.

RX PubMed=9660949;

RA Grueter P., Taberner C., von Kobbe C., Schmitt C., Saavedra C.,

RA Bachi A., Wilm M., Felber B.K., Izaurralde E.;

RT "TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export

RT from the nucleus.";

RL Mol. Cell 1:649-659(1998).

RN [8]

RP CHARACTERIZATION.

RX MEDLINE=21282872; PubMed=11259411;

RA Braun I.C., Herold A., Rode M., Conti E., Izaurralde E.;

RT "Overexpression of TAP/p15 heterodimers bypasses nuclear retention and

RT stimulates nuclear mRNA export.";

RL J. Biol. Chem. 276:20536-20543(2001).

RN [9]

RP CHARACTERIZATION.

RX PubMed=10668806;

RA Bachi A., Braun I.C., Rodrigues J.P., Pante N., Ribbeck K.,

RA von Kobbe C., Kutay U., Wilm M., Gorlich D., Carmo-Fonseca M.,

Izaurrealde E.;  
 "The C-terminal domain of TAP interacts with the nuclear pore complex  
 and promotes export of specific CTE-bearing RNA substrates.";  
 RNA 6:136-158(2000).  
 [10]  
 RA MUTAGENESIS.  
 RX MEDLINE=21151125; PubMed=11256625;  
 RA Suyama M., Doerks T., Braun I.C., Sattler M., Izaurralde E., Bork P.;  
 RA "Prediction of structural domains of TAP reveals details of its  
 RT interaction with p15 and nucleoporins.";  
 RL EMBO Rep. 1:53-58(2000).  
 [11]  
 RN X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 102-372.  
 RP MEDLINE=20514125; PubMed=11060011;  
 RA Liker E., Fernandez E., Izaurralde E., Conti E.;  
 RA "The structure of the mRNA export factor TAP reveals a cis arrangement  
 RT of a non-canonical RNP domain and an LRR domain.";  
 RL EMBO J. 19:5587-5598(2000).  
 [12]  
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH NXT1, AND X-RAY  
 RP CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF COMPLEX WITH NXT1-FG-REPEAT.  
 RX PubMed=11583626;  
 RA Fribourg S., Braun I.C., Izaurralde E., Conti E.;  
 RA "Structural basis for the recognition of a nucleoporin FG repeat by  
 RT the NTF2-like domain of the TAP/p15 mRNA nuclear export factor.";  
 RL Mol. Cell 8:645-656(2001).  
 [13]  
 RN STRUCTURE BY NMR OF 551-619, AND MUTAGENESIS OF PHE-617.  
 RP MEDLINE=21912422; PubMed=11875519;  
 RA Grant R.P., Hurt E., Neuhaus D., Stewart M.;  
 RA "Structure of the C-terminal FG-nucleoporin binding domain of  
 RT Tap/NXF1.";  
 RL Nat. Struct. Biol. 9:247-251(2002).  
 CC -!- FUNCTION: Involved in the nuclear export of mRNA species bearing  
 CC retroviral constitutive transport elements (CTE) and in the export  
 CC of mRNA from the nucleus to the cytoplasm.  
 CC -!- SUBUNIT: Interacts with NXT1, NXT2, EIB-AP5, RAE1, the REF  
 CC proteins and with several nucleoporins.  
 CC -!- SUBCELLULAR LOCATION: Nuclear; localized predominantly in the  
 CC nucleoplasm and at both the nucleoplasmic and cytoplasmic faces of  
 CC the nuclear pore complex. Shuttles between the nucleus and the  
 CC cytoplasm.  
 CC -!- TISSUE SPECIFICITY: Expressed ubiquitously.  
 CC -!- DOMAIN: The minimal CTE binding domain consists of an RNP-  
 CC type RNA binding domain (RBD) and leucine-rich repeats.  
 CC -!- DOMAIN: The nucleoporin binding domain consists of a NTF2-like  
 CC domain and a UBA-like domain. The NTF2 domain heterodimerizes with  
 CC for NXF1 AND NXT2. The formation of NXF1/NXT1 heterodimers is required  
 CC for NXF1-mediated nuclear mRNA export. The UBA-like domain  
 CC mediates direct interactions with nucleoporin-FG-repeats and is  
 CC necessary and sufficient for localization of NXF1 to the nuclear  
 CC rim. The conserved loop 594-NWD-596 of the UBA domain has a  
 CC critical role in the interaction with nucleoporins.  
 CC -!- DOMAIN: The leucine-rich repeats and the NTF2-domain are  
 CC essential for the export of mRNA from the nucleus.  
 CC -!- MISCELLANEOUS: The RNA-binding domain is a non-canonical RNP-type  
 CC domain.  
 CC -!- SIMILARITY: BELONGS TO THE NXF FAMILY.  
 CC -!- SIMILARITY: CONTAINS 4 LEUCINE-RICH REPEATS (LRR).  
 CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -!- SIMILARITY: CONTAINS 1 UBA DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AJ132712; CAA10753.1; -.  
 CC EMBL; AF12880; AAD39102.1; -.  
 DR

DR EMBL; AF126246; AAD20016.1; -.  
 DR EMBL; BC004904; AAO4904.1; -.  
 DR EMBL; AK027192; -; NOT ANNOTATED\_CDS.  
 DR EMBL; U80073; AAB8111.1; -.  
 DR DR Genew; HGNC:8071; NXF1.  
 DR MIM; 602647; -.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003603; LRRcap.  
 DR PDB; 1G05; 18-OCT-01.  
 DR PDB; 1JRG; 12-JUL-01.  
 DR PDB; 1JNS; 23-JUL-01.  
 DR Pfam; PF00560; LRR; 2.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR SMART; SM00446; LRRcap; 1.  
 DR PROSITE; PS0177; NTF2 DOMAIN; 1.  
 DR Transport; mRNA transport; Nuclear protein; RNA-binding; Repeat;  
 KW Leucine-rich repeat; Multigene family; 3D-structure.  
 FT DOMAIN 119 198 RNA-BINDING (RRM).  
 FT REPEAT 266 291 LRR 1.  
 FT REPEAT 292 315 LRR 2.  
 FT REPEAT 316 343 LRR 3.  
 FT REPEAT 344 371 LRR 4.  
 FT DOMAIN 386 536 NTF2.  
 FT DOMAIN 566 608 UBA-LIKE.  
 FT DOMAIN 67 100 NUCLEAR LOCALIZATION SIGNAL.  
 FT DOMAIN 83 110 NUCLEAR EXPORT SIGNAL.  
 FT DOMAIN 551 561 PRO-RICH.  
 FT MUTAGEN 306 308 ERE->AAA: DECREASES THE EXPORT OF MRNAS  
 FROM THE NUCLEUS.  
 FT MUTAGEN 594 594 W->A: SUPPRESSES FG-NUCLEOPORIN BINDING.  
 FT MUTAGEN 595 595 D->R: SUPPRESSES FG-NUCLEOPORIN BINDING.  
 FT MUTAGEN 617 617 F->A: SUPPRESSES FG-NUCLEOPORIN BINDING.  
 FT CONFLICT 119 119 W -> C (IN REF. 6).  
 FT CONFLICT 256 256 T -> N (IN REF. 3).  
 SQ SEQUENCE 619 AA; 70182 MW; 339872AADA789FBF CRC64;

Query Match 7.6%; Score 93; DB 1; Length 619;  
 Best Local Similarity 24.7%; Pred. No. 0.49;  
 Matches 60; Conservative 32; Mismatches 93; Indels 58; Gaps 13;

Qy 5 AGDSGML-----ACLTVLVHLPAVPAALNRTGDP-GPGPSIQKTYDILT---RYLEHQL 54  
 Db 400 SGDRGLLDVHDGACCSLSIPFQNPASRLAEYFKDSRVNKKLPTLRFLKHTR 459  
 Qy 55 RSLAATYLNLYGPPNEEDFNP--PRLGATLPRATVDL-----EVRSLNDKLR-LTON 106  
 Db 460 LNVV-AFLNEL--PKTDQVNSFVVDISAQTSLLCFSVNGVFKVDGKSRDSLRAFTRT 516  
 Qy 107 Y---RAYSHLLCYLRG--LNROAATAELRRSLAHFCTSLQGLLGSIAGVMAALGYPLPQP 161  
 Db 517 FIAPVPSNSGLCIVNDELFRVNASBEEIQRAFA-----MPAP 553  
 Qy 162 LPTGPTTPGPAHSDFLQK-----MDDFWLLKELQTLW---RSKADFNRLKKMQPP 212  
 Db 554 TPSSSPVPTLSPQQEMLQAFSTQSGNLEWSQKCLQDNNDWYTRSAQAQTHLKAKGEIP 613  
 Qy 213 AAA 215  
 Db 614 EVA 616

RESULT 5  
 CTF1 HUMAN STANDARD; PRT; 201 AA.  
 ID CTF1 HUMAN  
 AC Q16619;

DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cardiotoxin-1 (CT-1).  
 GN CTF1.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.





OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92304573; PubMed=1610564;  
 RA Leung D.W., Parent A.S., Cachianes G., Lee A.L., Nikolics K.,  
 RA Esch F., Coulombe J.N., Blacher R.W., Eckenstein F.P., Nishi R.;  
 RT "Cloning, expression during development, and evidence for release of  
 RT a trophic factor for ciliary ganglion neurons";  
 RL Neuron 8:1045-1053(1992).  
 CC -1- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL  
 CC AXOTOMY.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM.  
 CC -1- SIMILARITY: BELONGS TO THE CNTF FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M80827; AAA48784.1; -;  
 DR PIR; JH0680; JH0680.  
 DR HSSP; P26441; ICNT.  
 DR InterPro; IPR000151; Cil\_neuro\_factor.  
 DR Pfam; PF01110; CNTF; 1.  
 DR ProDom; PD011041; Cil\_neuro\_factor; 1.  
 DR Growth factor; Neurone.  
 KW Growth factor; Neurone.  
 SQ SEQUENCE 195 AA; 21330 MW; FEA076949DB34AC5 CRC64;  
 Query Match 7.4%; Score 91; DB 1; Length 195;  
 Best Local Similarity 27.3%; Pred. No. 0.18;  
 Matches 51; Conservative 21; Mismatches 83; Indels 32; Gaps 9;  
 QY 46 LRYLEHQLRSAGTYLVYLPFPNPPRLGAEPLPRATVLEWVRSINDKRLTLQ 105  
 DB 23 LARKMSRSDVTLLDIIVVERQG-----LDASISVAADVGVPTAAV--ERWAEQTGTQLLD 75  
 QY 106 N---YEAYSHLLCYLRGLNRQA---ATAELRSLA-----HFCTSLOGLLGSIGAWMA 152  
 DB 76 NLAAVFAFRTLQAQMLEEQRELLGDTDAELGPAALAMLLQSAFVYHLELL-----ELE 130  
 QY 153 ALGYPLPQPLGTEPTWTGPAH-SPLQKMDDFWLKELQTLWRSKADFNRLKKQKP 211  
 DB 131 SRGAPAE---GSEP---PAPPLSLFEQKLRGLRVLRLAQAQAVRSVRDLRLSKHGP 184  
 QY 212 PAAAVTL 218  
 DB 185 SGAALGL 191  
 RESULT 8  
 MIP1\_SCHPO STANDARD; PRT; 1313 AA.  
 AC P87141;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE WD-repeat protein mipl.  
 GN MIP1 OR SPAC57A7.11.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.; FUNCTION, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=20115869; PubMed=10648609;  
 RA Shinozaki-Yabana S., Watanabe Y., Yamamoto M.;

RT "Novel WD-repeat protein Miplp facilitates function of the meiotic  
 RT regulator Mei2p in fission yeast."; Mol. Cell. Biol. 20:1234-1242(2000).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- FUNCTION: BINDS TO AND FACILITATES THE FUNCTIONING OF THE MEIOTIC  
 CC REGULATOR MEI2. MAY ALSO BE INVOLVED IN CONJUGATION BY INTERACTING  
 CC WITH STE11. ESSENTIAL FOR CELL GROWTH.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB032552; BAA84595.1; -;  
 DR EMBL; Z95396; CAB08769.1; -;  
 DR InterPro; IPR001680; WD40.  
 DR InterPro; IPR004083; Yeast176.  
 DR Pfam; PF00400; WD40; 5.  
 DR PRINTS; PR00320; GPROTEINRPT.  
 DR PRINTS; PR01547; YEAST176DUF.  
 DR SMART; SM00320; WD40; 5.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 2.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Meiosis; WD repeat; Repeat.  
 FT REPEAT 986 1029 WD 1.  
 FT REPEAT 1033 1074 WD 2.  
 FT REPEAT 1087 1126 WD 3.  
 FT REPEAT 1130 1170 WD 4.  
 FT REPEAT 1176 1216 WD 5.  
 FT REPEAT 1219 1259 WD 6.  
 FT REPEAT 1268 1308 WD 7.  
 SQ SEQUENCE 1313 AA; 148533 MW; C71B663B0171E7A4 CRC64;  
 Query Match 7.3%; Score 89.5; DB 1; Length 1313;  
 Best Local Similarity 22.8%; Pred. No. 2.7;  
 Matches 61; Conservative 32; Mismatches 69; Indels 105; Gaps 14;  
 QY 15 LCTVLWHLPAV--PALNR-----TGDPGPGPSI-----QKTYDLTRYLEHQLRS 56

```
Db 811 LAFLLQHLPALHAKASLKSDTNSVTSDPKPHFVPSVSENKILNRSFSLRSLKGLALS 870
QY 57 LAG-----TVLNYL-----GPP-FNEPDFNPPRLGAETLPRATVDL 91
Db 871 LAGSDRASELLSNGENKPAESNLNHLTSKVPFPFAFNELEY-----QSELDW 919
QY 92 EVWRSND-----KRLTQNYEAYSHLLCYLRGL-----NRQAATAELRRSLAHFCTS 139
Db 920 PLTSYLFDSRSKYFTEPQMRPNDEPGS--ICYNQLWRNERNEXLIYTRPLAEYSTN 977
QY 140 -----LQGLLSGIAGVMAALGVPLOPLGTEPTWPGPAHSDFLQKMDDFWLKKE--- 190
Db 978 GRWNOQLMTFNNTIA-----PRKLMFHQFEDQLITLGDKDI 1013
QY 191 LQTLWLRSAKDNRLKKMKOPPPAAAVT 217
Db 1014 IQVDWRR-----NRLNSFKTSASAT 1036
RESULT 9
HAIR_RAT
ID HAIR_RAT STANDARD; PRT; 1181 AA.
AC F97609;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Hairless protein.
GN HR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=97141510; PubMed=8987811;
RA Thompson C.C.;
RT "Thyroid hormone-responsive genes in developing cerebellum include a
RT novel synaptotagmin and a hairless homolog.";
RL J. Neurosci. 16:7832-7840(1996).
CC -!- FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR THAT COULD ACT ON TO
CC REGULATE ONE OF THE PHASES OF HAIR GROWTH.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U71293; AAC3018.1; ALT INIT.
DR InterPro; IPR003347; TF_JmJC.
DR Pfam; PF02373; jmjC; 1.
KW Zinc-finger; DNA-binding; Nuclear protein; Transcription regulation;
KW Metal-binding.
FT ZN FING 594 619 C6-TYPE.
SQ SEQUENCE 1181 AA; 127307 MW; 834B7029CF8E88F0 CRC64;
Query Match 7.1%; Score 87.5; DB 1; Length 1181;
Best Local Similarity 25.3%; Pred. No. 3.6;
Matches 61; Conservative 21; Mismatches 92; Indels 67; Gaps 12;
QY 30 RTGDPGSGPSIQKTYDITRYLHQLRSLAGTY--LNVIGPP-----FNEPDF 74
Db 859 RQGPVLVSGTQKTLRLSLMGWELTGLGGVQTLTALGPPQPTSLDSTAFKGFSPHEA 918
QY 75 NPPELGAETLPRATVDLEVRSLNDK-----LRLTQ-----NREAYSHLIC 115
Db 919 RP-----KLDEGSV-LLLHRLPDKGDESRENVLASSLPLPEYCAHQKUNLASLPLGL 971
```

```
QY 116 YLRGLNRQAATAELRRS-----LAHFTSLOGLLSIAGVMAALGYPLPQPLGTEPTW 169
Db 972 TLHLEPQLWAAVGNVSHRGLGTKNLCVEVSDLSILVHAEAL-----PPW 1019
QY 170 TPCPAHSDFLQKMD--DFWLLKELQTLWR--SAKDFNRLKKMKOP--PAAAVTLHLGAH 223
Db 1020 Y--RAQKDFSLGDLGEGSLWSPGSGTSTVHVHVFRAQDAQRIRFLQWVCPAGAGTLBPGAP 1077
QY 224 G 224
Db 1078 G 1078
RESULT 10
M2B2 HUMAN
ID M2B2 HUMAN STANDARD; PRT; 1009 AA.
AC Q9Y2E5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epididymis-specific alpha-mannosidase precursor (EC 3.2.1.24)
DE (Mannosidase alpha class 2B member 2).
GN MAN2B2 OR KIAA0935.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Stone N.E., Schmutz J.J., Cox D.R., Myers R.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 132-1009 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
RN [3]
RP RECONSTRUCTION FROM GENOMIC SEQUENCE.
RA Bairoch A.;
RL Unpublished observations (NOV-2001).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-
CC mannose residues in alpha-D-mannosides.
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC004480; -- NOT ANNOTATED CDS.
DR EMBL; AB023152; BAA76779.1; ALT_SEQ.
KW Hydrolase; Glycosidase; Signal; Glycoprotein.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1009 EPIDIDYMIS-SPECIFIC ALPHA-MANNOSIDASE.
FT CARBOHYD 226 226 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 516 516 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 608 608 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 675 675 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 808 808 N-LINKED (GLCNAC. .) (POTENTIAL).
```



```

Db 128 VLEHKKVPPSADGTPLSVGGGLFEKKLWGLKVLQELSQWTVSRDLRVISHSQ---- 183
QY 214 AAVTLHLGAAH 223
Db 184 AGVPAH-GSH 192

RESULT 13
TLL_DROME
ID TLL_DROME STANDARD; PRT; 452 AA.
AC P18102; Q9VA33;
DT 01-NOV-1990 (Rel. 16, Created)
DT 16-OCT-2001 (Rel. 16, Last sequence update)
DE Tailless protein.
GN TLL OR NR2E2 OR CG1378.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=90304905; PubMed=2364433;
RA Pignoni F., Baldarelli R.M., Steingrimsson E., Diaz R.J.,
RA Patapoutian A., Merriam J.R., Lengyel J.A.;
RT "The Drosophila gene tailless is expressed at the embryonic termini
RT and is a member of the steroid receptor superfamily.";
RL Cell 62:151-163 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93157371; PubMed=8430097;
RA Liaw G.-J., Steingrimsson E., Pignoni F., Courey A.J., Lengyel J.A.;
RT "Characterization of downstream elements in a Raf-1 pathway.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:858-862 (1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer J., Brokstein P., Brottier P.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Arlil J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskaas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

```

```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [4]
RP FUNCTION.
RX MEDLINE=99287803; PubMed=10357938;
RA Daniel A., Dumstrei K., Lengyel J.A., Hartenstein V.;
RT "The control of cell fate in the embryonic visual system by atonal,
RT tailless and EGFR signaling.";
RL Development 126:2945-2954 (1999).
CC -|- FUNCTION: ORPHAN RECEPTOR THAT BINDS DNA AS A MONOMER TO HORMONE
CC RESPONSE ELEMENTS (HRE) CONTAINING AN EXTENDED CORE MOTIF HALF-
CC SITE SEQUENCE 5'-AAGTCA-3' IN WHICH THE 5' FLANKING NUCLEOTIDES
CC PARTICIPATE IN DETERMINING RECEPTOR SPECIFICITY. THIS RECEPTOR
CC BINDS TO THE CONSENSUS SEQUENCE [AG][AG]AAGTCA. PLAYS A KEY ROLE
CC IN THE ESTABLISHMENT OF NONMETAMERIC DOMAINS AT THE ANTERIOR AND
CC POSTERIOR POLES OF THE EMBRYO. IT MAY ALSO PLAY A ROLE IN THE
CC NERVOUS SYSTEM. THE MATERNAL TERMINAL PATHWAY ACTIVATES THE TLL
CC GENE IN THE TERMINI; TLL ACTIVITY THEN REPRESSES SEGMENTATION AND
CC ACTIVATES TERMINAL-SPECIFIC GENES IN THESE DOMAINS. INVOLVED IN
CC THE REGULATION OF EARLY EYE DEVELOPMENT. IN THE EMBRYONIC VISUAL
CC SYSTEM ANLAGE DRIVES CELLS TO OPTIC LOBE AS OPPOSED TO BOLWIG'S
CC ORGAN FATE.
CC -|- SUBUNIT: MONOMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -|- TISSUE SPECIFICITY: BRAIN AND PERIPHERAL NERVOUS SYSTEM.
CC -|- DEVELOPMENTAL STAGE: DURING STAGE 10 FOUND IN THE ANTERIOR PART OF
CC THE VISUAL SYSTEM THAT LATER GIVES RISE TO THE ANTERIOR LIP OF THE
CC OPTIC LOBE. AT STAGE 12 ALSO FOUND IN THE POSTERIOR LIP OF THE
CC OPTIC LOBE. IN THIRD LARVAL INSTAR EXPRESSED IN THE OPTIC LOBE OF
CC THE LARVAL BRAIN AND IN THE EYE ANTENNAL DISK, BOTH IN ANTENNAL
CC AND EYE PORTION.
CC -|- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR2 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M34639; AAA28936.1; -.
CC EMBL; AF019362; AAF71371.1; -.
CC EMBL; AE003775; AAF57091.1; -.
CC PIR; A35602; A35602.
CC PIR; A47265; A47265.
CC HSPSP; P10826; JHRA.
CC TRANSFAC; T00789; -.
CC FlyBase; FBgn0003720; tll.
CC InterPro; IPR000536; Hormone_rec_lig.
CC InterPro; IPR001628; Znf_C4steroid.
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF00105; zf_C4; 1.
CC PRINTS; PR00047; STROIDFINGER.
CC ProDom; PD000035; Znf_C4steroid; 1.
CC SMART; SM00430; HOL1; 1.
CC SMART; SM00399; Znf_C4; 1.
CC PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;
CC Zinc-finger; Activator; Repressor; Developmental protein.
CC DNA_BIND 34 101 NUCLEAR RECEPTOR-TYPE.
CC ZN_FING 34 54 C4-TYPE.
CC ZN_FING 70 96 C4-TYPE.
CC DOMAIN 244 389 LIGAND-BINDING (BY SIMILARITY).
CC DOMAIN 262 265 POLY-LEU.
CC SEQUENCE 452 AA; 50549 MW; A4ABEFFDE993A37C CRC64;
SQ

```

Query Match

6.8%; Score 83; DB 1; Length 452;

```
Best Local Similarity 23.8%; Pred. No. 2.9;
Matches 34; Conservative 11; Mismatches 32; Indels 66; Gaps 6;

Qy 115 CYLRGLNRQAATAE-----LRRSLAHFCTSLQGLSLAGVW-----AALG 155
Db 96 CFEVGNKDAVQHERGPRNSTRHRWAMYKDAWVG-----AGEMPQIPAEILMNTAALTG 150
Qy 156 YP-LPQPLFGTFTPTGPAHSDFLQMDDFWLLKELQTLWMSAKDFNRLKKRMQPPAA 214
Db 151 PFGVPMPEGLPQRAGHHPAHMAAFQ-----PPPSA 181
Qy 215 AVTL-----HLGAHF 225
Db 182 AAVLDLSVRVPHHPVHQHGF 204

RESULT 14
HAIR_HUMAN
ID HAIR_HUMAN STANDARD; PRT; 1189 AA.
AC O43593; Q9NPEI;
DT 15-JUL-1999 (Rel. 38, "Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hairless protein.
GN HR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM), AND VARIANT ALUNC ALA-1022.
RX MEDLINE=98111413; PubMed=9445480;
RA Ahmad W., ul Haque M.F., Brancolini V., Tsou H.C., Ul Haque S.,
RA Lam H., Alta V.M., Owen J., Deblaquiere M., Frank J.,
RA Cserhalmi-Friedman P.B., Leask A., McGrath J.A., Peacocke M.,
RA Ahmad M., Ott J., Christiano A.M.;
RA "Alopecia universalis associated with a mutation in the human hairless
RT gene.";
RT Science 279:720-724(1998).
RL [2]
RP SEQUENCE FROM N.A. (LONG ISOFORM), REVISIONS TO 572 AND 774, AND
RP TISSUE SPECIFICITY.
RP TISSUE SPECIFICITY.
RC TISSUE=Peripheral blood leukocytes, and Skin fibroblast;
RX MEDLINE=99162400; PubMed=10051399;
RA Ahmad W., Zlotogorski A., Panteleyev A.A., Lam H., Ahmad M.,
RA ul Haque M.F., Abdallah H.M., Dragan L., Christiano A.M.;
RA "Genomic organization of the human hairless gene (HR) and
RT identification of a mutation underlying congenital atrichia in an Arab
RT Palestinian family.";
RT Genomics 56:141-148(1999).
RL [3]
RP SEQUENCE FROM N.A., VARIANT ALUNC ASP-1136, TISSUE SPECIFICITY, AND
RP ALTERNATIVE SPLICING.
RP TISSUE=Peripheral blood leukocytes, Brain, and Fetal brain;
RX MEDLINE=98409496; PubMed=9736769;
RA Cichon S., Anker M., Vogt I.R., Rohleder H., Putzstuck M., Hillmer A.,
RA Farooq S.A., Al-Dhafri K.S., Ahmad M., Haque S., Rietschel M.,
RA Parrooping P., Kruse R., Noeche M.M.;
RA "Cloning, genomic organization, alternative transcripts and mutational
RT analysis of the gene responsible for autosomal recessive universal
RT congenital alopecia.";
RL Hum. Mol. Genet. 7:1671-1679(1998).
RL [4]
RP VARIANT APL GLN-620.
RX MEDLINE=98431781; PubMed=9758627;
RA Ahmad W., Irvine A.D., Lam H., Buckley C., Bingham E.A.,
RA Panteleyev A.A., Ahmad M., McGrath J.A., Christiano A.M.;
RT "A missense mutation in the zinc-finger domain of the human hairless
RT gene underlies congenital atrichia in a family of Irish travellers.";
RL Am. J. Hum. Genet. 63:984-991(1998).
CC -!- FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR THAT COULD ACT ON TO
CC -!- REGULATE ONE OF THE PHASES OF HAIR GROWTH.
CC -!- SUBCELLULAR LOCATION: Nuclear.
```

```
CC -!- ALTERNATIVE PRODUCTS: At least 2 isoforms; a long form (shown
CC here) and a short form; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Strongest expression of both isoforms is seen
CC in the small intestine, weaker expression in brain and colon, and
CC trace expression is found in liver, pancreas, spleen, thymus,
CC stomach, salivary gland, appendix and trachea. Long isoform is
CC always the most abundant. Long isoform is exclusively expressed at
CC low levels in kidney and testis and short isoform exclusively at
CC high levels in the skin.
CC -!- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ALOPECIA UNIVERSALIS
CC (ALUNC); A RARE AUTOSOMAL RECESSIVE FORM OF HAIR LOSS. IT IS
CC CHARACTERIZED BY HAIR FOLLICLES WITHOUT HAIR.
CC -!- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ATRICHIA WITH PAPULAR
CC LESIONS (APL) (ALSO KNOWN AS CONGENITAL ATRICHIA). THIS AUTOSOMAL
CC RECESSIVE DISEASE IS CHARACTERIZED BY PAPILLARY LESIONS OVER MOST
CC OF THE BODY AND ALMOST COMPLETE ABSENCE OF HAIR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF039196; AAC32258.2; -.
CC EMBL; AJ277249; CAB87577.2; -.
CC EMBL; AJ277250; CAB87577.2; JOINED.
CC EMBL; AJ277251; CAB87577.2; JOINED.
CC EMBL; AJ277252; CAB87577.2; JOINED.
CC EMBL; AJ277253; CAB87577.2; JOINED.
CC EMBL; AJ400825; CAB87577.2; JOINED.
CC EMBL; AJ400826; CAB87577.2; JOINED.
CC EMBL; AJ400827; CAB87577.2; JOINED.
CC EMBL; AJ400828; CAB87577.2; JOINED.
CC EMBL; AJ400829; CAB87577.2; JOINED.
CC EMBL; AJ400830; CAB87577.2; JOINED.
CC EMBL; AJ400831; CAB87577.2; JOINED.
CC EMBL; AJ400832; CAB87577.2; JOINED.
CC EMBL; AJ400833; CAB87577.2; JOINED.
CC EMBL; AJ400834; CAB87577.2; JOINED.
CC EMBL; AJ400835; CAB87577.2; JOINED.
CC EMBL; AJ400836; CAB87577.2; JOINED.
CC EMBL; AJ400837; CAB87577.2; JOINED.
CC EMBL; AJ277165; CAB86602.1; -.
CC EMBL; HGNC:5172; HR.
CC EMBL; 602302; -.
CC EMBL; 203655; -.
CC EMBL; 209500; -.
CC EMBL; InterPro; IPR003347; TF_JmjC.
CC EMBL; Pfam; PF02373; jmjC; 1.
CC EMBL; Zinc-finger; DNA-binding; Nuclear protein; Transcription regulation;
CC Metal-binding; Alternative splicing; Disease mutation.
CC EMBL; C6-TYPE. 625
CC EMBL; ZNFING 600 625
CC EMBL; VARSPLIC 1072 1126
CC EMBL; VARIANT 620 620
CC EMBL; R -> Q (IN APL).
CC EMBL; /FTid=VAR_005265.
CC EMBL; T -> A (IN ALUNC).
CC EMBL; /FTid=VAR_005266.
CC EMBL; V -> D (IN ALUNC).
CC EMBL; /FTid=VAR_005267.
CC EMBL; D -> G (IN REF. 3).
CC EMBL; L -> W (IN REF. 3).
CC EMBL; S -> G (IN REF. 3).
CC EMBL; CONFLICT 337 337
CC EMBL; CONFLICT 446 446
CC EMBL; CONFLICT 584 584
CC EMBL; CONFLICT 1189 AA; 127509 MW; 5E244858716EB5DF CRC64;
CC EMBL; SEQUENCE 1189 AA; 127509 MW; 5E244858716EB5DF CRC64;

Query Match 6.7%; Score 82.5; DB 1; Length 1189;
Best Local Similarity 23.3%; Pred. No. 11;
Matches 71; Conservative 23; Mismatches 92; Indels 119; Gaps 15;

Qy 3 LRAGDSW-GMLACLCVTLVHLPAVPAALNRGTGDPGP-----SI 40
Db 818 LRAGPGLRKGGLPLSPVRPLPPPPGALLWLQEPQPCPRGPHLFQEHWRQGPVLVSGI 877
```







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:17 ; Search time 33 Seconds  
(without alignments)  
908.527 Million cell updates/sec

Title: US-09-931-704-2

Perfect score: 1226

Sequence: 1 MDLRAGDSWGLACTLTVL.....KKMQPPAAAVTLHGAGHF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A. Geneseq 101002.\*

1: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SID82/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SID82/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SID82/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1226	100.0	225	19 AAW29715	Human neurotrophic
2	1226	100.0	225	19 AAW56141	Amino acid sequenc
3	1226	100.0	225	20 AAW94466	Human cardiotothro
4	1226	100.0	225	21 AAY87813	Human NNT-1 protei
5	1226	100.0	225	22 AAG63543	Amino acid sequenc
6	1226	100.0	225	23 AAU78176	Human novel neurot
7	1226	100.0	225	22 AAM25831	Human protein sequ
8	1214	99.0	223	22 AAE00828	Human cardiotothro
9	1204	98.2	321	22 ABB11896	Human cardiotothro
10	1204	98.2	321	22 AAM79399	Human protein SEQ

11	1197	97.6	260	22 AAM78415	Human protein SEQ
12	1193	97.3	225	19 AAW29716	Mouse neurotrophic
13	1193	97.3	225	19 AAW56142	Amino acid sequenc
14	1193	97.3	225	21 AAY87814	Murine NNT-1 prote
15	1193	97.3	225	23 AAU78177	Mouse novel neurot
16	1169	95.4	215	21 AAB19586	Human interleukin-
17	1136	92.7	215	21 AAB19587	Mouse interleukin-
18	885	72.2	164	22 ABB40317	Peptide #7823 enco
19	885	72.2	164	22 ABB24716	Protein #6715 enco
20	885	72.2	164	22 AAM61118	Human brain expres
21	885	72.2	164	22 AAM73827	Human bone marrow
22	885	72.2	164	22 AAM20115	Peptide #6549 enco
23	885	72.2	164	22 AAM34012	Peptide #8049 enco
24	885	72.2	164	23 AAG43716	Human DNAX interle
25	162.5	13.3	208	20 AAY09197	Human DNAX interle
26	160	13.1	208	20 AAY09196	Human DNAX interle
27	118.5	9.7	203	16 AAR83965	Mouse cardiac hype
28	118.5	9.7	203	17 AAR88204	Human cardiotothro
29	118.5	9.7	203	18 AAR29237	Murine cardiotothro
30	96.5	7.9	243	22 AAU09153	Human cytokine Zal
31	96.5	7.9	243	22 AAB20277	Human interleukin
32	96.5	7.9	243	23 AAU76375	Human helical prot
33	92	7.5	332	21 AAG22132	Arabidopsis thalia
34	92	7.5	332	21 AAG40321	Arabidopsis thalia
35	91.5	7.5	201	16 AAR83967	Human cardiac hype
36	91.5	7.5	201	18 AAW29238	Human cardiotothro
37	91.5	7.5	201	20 AAY06490	Human tumour-associ
38	91.5	7.5	201	21 AAB27662	Human protein PRO8
39	91.5	7.5	201	21 AAB13004	Human cardiotothro
40	91.5	7.5	201	21 AAY93697	Amino acid sequenc
41	91.5	7.5	201	21 AAY87818	Human cardiotothro
42	91.5	7.5	201	22 AAB50994	Human PRO882 prote
43	91.5	7.5	1182	23 AAE19798	Mouse Hairless pro
44	91	7.4	195	14 AAR34432	Sequence of growth
45	91	7.4	195	20 AAR83337	Chicken ciliary ne

#### ALIGNMENTS

RESULT 1  
AAW29715  
ID AAW29715 standard; Protein; 225 AA.  
XX  
AC AAW29715;  
XX  
DT 09-NOV-1998 (first entry)  
XX  
DE Human neurotrophic factor NNT-1.

XX NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
KW peripheral neuropathy; dystrophy; neural retina degeneration;  
KW common variable immunodeficiency; CVID; selective IGA deficiency;  
KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;  
XX therapy.  
XX  
OS Homo sapiens.

XX	Key	Location/Qualifiers
FT	Peptide	1..27
FT	Protein	/label= sig_peptide
FT		28..225
FT		/label= Mat_protein
XX		WO9833922-A1.
XX		06-AUG-1998.
XX		02-FEB-1998; 98WO-US02363.
XX		30-JAN-1998; 98US-0016534.

```

PR 03-FEB-1997; 97US-0792019.
XX (AMGE-) AMGEN INC.
PA
XX Chang M, Elliot GS, Sarmiento U, Senaldi G;
XX N-PSDB; AAV47510-11.
XX WPI; 1998-437475/37.
XX DR N-PSDB; AAV47510-11.
XX
XX Newly isolated nucleic acid encoding human or murine neurotrophic
PT factor NNT-1 - useful for treatment of neurological and
PT immunological diseases or inflammation, also as vaccine adjuvant
XX
XX Claim 12; Fig 3; 120pp; English.
XX
XX This is the amino acid sequence of a novel neurotrophic factor,
CC designated NNT-1, that is a growth factor for neurons and for B or
CC T cells. It was deduced from isolated cDNA (see AAV47510) and
CC genomic DNA (see AAV47511) clones. Vectors containing the cDNA or
CC genomic DNA and host cells are provided for use in the production
CC of NNT-1 polypeptides. These are used to treat: (i) neurological
CC or immunological diseases, specifically Alzheimer's, Parkinson's
CC or Huntington's diseases, amyotrophic lateral sclerosis,
CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and
CC degeneration of the neural retina, or conditions characterised by T
CC or B cell defects, e.g. common variable immunodeficiency (CVID),
CC selective IgA deficiency, hypogammaglobulinaemia and X-linked
CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)
CC inflammation. NNT-1 is also able to boost immunoreactivity and
CC antibody production following vaccination, and, since it inhibits
CC tumour necrosis factor production, it may also be useful for
CC treating sepsis. In addition, cells that have been engineered to
CC express NNT-1 can be implanted, or nucleic acids are delivered in
CC gene therapy vectors.
XX
XX SQ Sequence 225 AA;
Query Match 100.0%; Score 1226; DB 19; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.4e-118;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGMGLACTVWLHPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGMGLACTVWLHPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
QY 61 YLNYLGPPNPDNPPRLGAETLPRAVDLEVWRSNDKRLTONYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPNPDNPPRLGAETLPRAVDLEVWRSNDKRLTONYEAYSHLLCYLRGL 120
QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTGPAHSDFLQ 180
Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTGPAHSDFLQ 180
QY 181 KMDDFWLLKELOTWLRSAKDFNRLKKMQPPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELOTWLRSAKDFNRLKKMQPPAAAVTLHLGAHGF 225

RESULT 2
AAW56141
ID AAW56141 standard; Protein; 225 AA.
XX
XX AAW56141;
XX
XX 13-JUL-1998 (first entry)
XX
XX Amino acid sequence of human neurotrophic factor NNT-1.
XX Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
XX treatment; neurological disease; degeneration; Parkinson's disease;
XX amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.
XX Homo sapiens.
XX

PR 03-FEB-1997; 97US-0792019.
XX (AMGE-) AMGEN INC.
PA
XX Chang M, Elliot GS, Sarmiento U, Senaldi G;
XX N-PSDB; AAV47510-11.
XX WPI; 1998-437475/37.
XX DR N-PSDB; AAV47510-11.
XX
XX Newly isolated nucleic acid encoding human or murine neurotrophic
PT factor NNT-1 - useful for treatment of neurological and
PT immunological diseases or inflammation, also as vaccine adjuvant
XX
XX Claim 12; Fig 3; 120pp; English.
XX
XX This is the amino acid sequence of a novel neurotrophic factor,
CC designated NNT-1, that is a growth factor for neurons and for B or
CC T cells. It was deduced from isolated cDNA (see AAV47510) and
CC genomic DNA (see AAV47511) clones. Vectors containing the cDNA or
CC genomic DNA and host cells are provided for use in the production
CC of NNT-1 polypeptides. These are used to treat: (i) neurological
CC or immunological diseases, specifically Alzheimer's, Parkinson's
CC or Huntington's diseases, amyotrophic lateral sclerosis,
CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and
CC degeneration of the neural retina, or conditions characterised by T
CC or B cell defects, e.g. common variable immunodeficiency (CVID),
CC selective IgA deficiency, hypogammaglobulinaemia and X-linked
CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)
CC inflammation. NNT-1 is also able to boost immunoreactivity and
CC antibody production following vaccination, and, since it inhibits
CC tumour necrosis factor production, it may also be useful for
CC treating sepsis. In addition, cells that have been engineered to
CC express NNT-1 can be implanted, or nucleic acids are delivered in
CC gene therapy vectors.
XX
XX SQ Sequence 225 AA;
Query Match 100.0%; Score 1226; DB 19; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.4e-118;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGMGLACTVWLHPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGMGLACTVWLHPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
QY 61 YLNYLGPPNPDNPPRLGAETLPRAVDLEVWRSNDKRLTONYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPNPDNPPRLGAETLPRAVDLEVWRSNDKRLTONYEAYSHLLCYLRGL 120
QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTGPAHSDFLQ 180
Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTGPAHSDFLQ 180
QY 181 KMDDFWLLKELOTWLRSAKDFNRLKKMQPPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELOTWLRSAKDFNRLKKMQPPAAAVTLHLGAHGF 225

RESULT 3
AAW94466
ID AAW94466 standard; Protein; 225 AA.
XX
XX AAW94466;
XX
XX 22-APR-1999 (first entry)
XX
XX Human cardiotrophin-like cytokine protein.
XX Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;
XX CLC; IL-6; diagnosis; detection; immune system-related disorder;
XX cancer; cardiac disorder; heart failure; hypertension; cancer;
XX autoimmune disorder; infection.
XX

```

```

OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..27
FT Protein 28..225
FT Domain 74..79
FT /label= Cardiotrophin-like_cytokine
FT /label= CD-I
FT /note= "conserved domain"
FT Domain 150..156
FT /label= CD-II
FT /note= "conserved domain"
FT Domain 194..198
FT /label= CD-III
FT /note= "conserved domain"
XX
PN WO9900415-A1.
XX
PD 07-JAN-1999.
XX
PF 29-JUN-1998; 98WO-US13129.
XX
PR 30-JUN-1997; 97US-0051311.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Shi Y;
XX
DR WPI; 1999-095678/08.
DR N-PSDB; AAX16161.
XX
PT New isolated cardiotrophin-like cytokine nucleic acid - used to
PT develop products for treating cardiac and immune system disorders,
PT e.g. heart failure, hypertension, cancers, autoimmune disorders and
PT infections
XX
PS Claim 1; Fig 1; 103pp; English.
XX
CC The present invention relates to a novel cardiotrophin-like cytokine
CC (CLC) protein which is a member of the interleukin 6 (IL-6) cytokine
CC family. The present sequence represents the human CLC protein. The
CC present invention also describes screening methods for identifying
CC agonists and antagonists of CLC activity, as well as methods for
CC detecting cardiac and immune system-related disorders and
CC therapeutic methods for treating cardiac and immune system-related
CC disorders, e.g. heart failure, hypertension, cancers, autoimmune
CC disorders and infections.
XX
SQ Sequence 225 AA;
Query Match 100.0%; Score 1226; DB 20; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.4e-118;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MDLRAGDSWGMCLACTVWLHPVAPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGMCLACTVWLHPVAPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
Qy 61 YLNYLGPPFPNEPDPNPRLGAEITLPRATVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPFPNEPDPNPRLGAEITLPRATVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120
Qy 121 NQQAATSLRSLAHFCTSLQGLGSIAGVMAALGYPIQPIPGTEPTWTGPAHSDFLQ 180
Db 121 NQQAATSLRSLAHFCTSLQGLGSIAGVMAALGYPIQPIPGTEPTWTGPAHSDFLQ 180
Qy 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAAAVTLHLGAHGF 225
RESULT 4

```

---

```

AA87813
ID AAY87813 standard; Protein; 225 AA.
XX
AC AAY87813;
XX
DT 24-AUG-2000 (first entry)
XX
DE Human NNT-1 protein.
XX
KW NNT-1; human; neurotrophic factor; neurotropic; neuroprotective; treatment;
KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
KW Huntington's disease; peripheral neuropathy; neural retina degeneration;
KW retinopathy; immune disorder; hematopoietic disorder.
XX
OS Homo sapiens.
XX
PN US6054294-A.
XX
PD 25-APR-2000.
XX
PF 12-DEC-1997; 97US-0988819.
XX
PR 03-FEB-1997; 97US-0792019.
XX
PA (AMGE-) AMGEN INC.
XX
PI Chang M;
XX
DR WPI; 2000-338492/29.
DR N-PSDB; AAA39481.
XX
PT New nucleic acids encoding neurotrophic factors useful for stimulating
PT growth of motor or sympathetic neurons for treating neuron cell damage
XX
PS Claim 1c; Fig 3; 42pp; English.
XX
CC This invention describes a novel nucleic acid molecule (I) encoding a
CC novel neurotrophic factor (NNT-1) (II) which has neurotropic,
CC neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and
CC ophthalmological activity. (I) is useful for producing NNT-1
CC polypeptides which are useful for treating patients in whom various
CC cells of the central, autonomic, or peripheral nervous system have
CC degenerated and/or have been damaged by congenital disease, trauma,
CC mechanical damage, surgery, stroke, ischemia, infection, metabolic
CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1
CC proteins are used to treat diseases like Alzheimer's, Parkinson's,
CC amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's
CC disease, peripheral neuropathy induced by diabetes or other metabolic
CC disorders, and/or dystrophies or degeneration of the neural retina such
CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of
CC night blindness, progressive cone-rod degeneration, immune disorders and
CC hematopoietic disorders. (I) is effective in treating neurological
CC conditions and promotes neuron regeneration. Neural functions are
CC effectively restored in patients suffering from various neurological
CC disorders. This sequence represents the human NNT-1 protein described in
CC the method of the invention.
XX
SQ Sequence 225 AA;
Query Match 100.0%; Score 1226; DB 21; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.4e-118;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MDLRAGDSWGMCLACTVWLHPVAPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGMCLACTVWLHPVAPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
Qy 61 YLNYLGPPFPNEPDPNPRLGAEITLPRATVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPFPNEPDPNPRLGAEITLPRATVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120

```

121	NRQAATAELRRSLAHFCTSLQGLIGSTAGVMAALGYPLPQPLPGTEPTWTTPGPAHSDFLQ	180
121	NRQAATAELRRSLAHFCTSLQGLIGSTAGVMAALGYPLPQPLPGTEPTWTTPGPAHSDFLQ	180
181	KMDDFWLLKELOTWLRSAKDFNRLKKKMOPPAAVTLHLGAGHF	225
181	KMDDFWLLKELOTWLRSAKDFNRLKKKMOPPAAVTLHLGAGHF	225
RESULT 5		
AAAG63543	AAG63543 standard; Protein; 225 AA.	
XX	AAAG63543;	
XX	15-OCT-2001 (first entry)	
XX	Amino acid sequence of a human NNT-1 protein.	
XX	NNT-1; CLF-1; SCNTFRalpha; nervous system; neuron; nervous system;	
XX	neuro-muscular function; tumour; immune system; haematopoietic system;	
XX	reproductive system; liver; skeletal muscle; neurodegenerative disease;	
XX	amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;	
XX	muscular mass; paralysis; cancer; obesity; fertility; endometriosis;	
XX	blastocyst implantation; thrombosis; retinal disease;	
XX	retinal pigmentosis.	
XX	Homo sapiens.	
XX	WO200155172-A2.	
XX	02-AUG-2001.	
XX	26-JAN-2001; 2001WO-FR00253.	
XX	27-JAN-2000; 2000FR-0001035.	
XX	12-OCT-2000; 2000FR-0013089.	
XX	(FABR ) FABRE MEDICAMENT SA PIERRE.	
XX	(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.	
XX	Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;	
XX	WPI; 2001-488773/53.	
XX	N-PSDB; AAH74484.	
XX	Claim 2; Page 58; 67pp; French.	
XX	The present sequence represents a human NNT-1 protein. The specification	
XX	describes a complex comprising a NNT-1 protein and a CLF-1 and/or	
XX	SCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate	
XX	activity of the SCNTFRalpha/gp130/LiFRbeta receptor complex, or to	
XX	induce phosphorylation of the tyrosine of gp130 and LiFRbeta, in the	
XX	particular where cells expressing the receptor complex are, in the	
XX	central or peripheral nervous system, in neurons implicated in	
XX	neuro-muscular function or in skeletal muscle. The complex or	
XX	antibodies are also used to decrease the survival, growth or	
XX	proliferation of tumour cells or to facilitate the proliferation and/or	
XX	inhibit differentiation of cells stocks. The complex is also used to	
XX	modulate activity of the gp130/LiFRbeta receptor or cells expressing	
XX	that receptor, particularly those cells implicated in the immune,	
XX	haematopoietic, nervous or reproductive system, the liver or skeletal	
XX	muscle. Molecules of the invention may be used to prevent or treat	
XX	neurodegenerative diseases including amyotrophic lateral sclerosis,	
XX	Parkinson's and Huntington's disease, to repair or regenerate nervous	
XX	or muscular tissue or to maintain muscular mass in paralysis patients.	
XX	They may also be used to treat cancer, obesity and associated diseases,	
XX	and to improve fertility, particularly to avoid endometriosis and/or	

inhibitor to a patient. Also included are a method of diagnosing an Igs-related disease or susceptibility to an IGE-related disease, by determining the presence or amount of expression of an NNT1 polypeptide encoded by a NNT1 nucleotide sequence, its fragment or naturally occurring variant, and diagnosing an Igs-related disease or susceptibility of an Igs-related disease based on the presence or amount of expression of the polypeptide and a pharmaceutical composition for use in treating IGE-related disease, comprising the NNT1 inhibitor. The NNT1 inhibitor is useful for preventing and treating IGE-related disease, modulating IGE levels, and treating allergic diseases e.g. Type I allergic disease, allergic rhinitis, eczema, dermatitis, pollinosis, asthma, immune diseases and disorders, diseases involving abnormal cell proliferation including cancer, arteriosclerosis and vascular stenosis, diseases and conditions relating to dysfunction of immune system including rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, osteoarthritis, inflammatory joint disease, autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory bowel disease, transplant rejection, and graft versus host disease, and reproductive diseases and disorders including infertility, miscarriage, preterm labour and delivery, and endometriosis. The present sequence represents human NNT1.

Query Match	100.0%	Score 1226;	DB 23;	Length 225;
Best Local Similarity	100.0%;	Pred. No. 1.4e-118;		
Matches 225;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MDLRAGDSWGMACLCCTVWLHPVAPALNRTGDPGPGPSIKQTYDLTRYLEHQRLSLAGT	60	
Db	1	MDLRAGDSWGMACLCCTVWLHPVAPALNRTGDPGPGPSIKQTYDLTRYLEHQRLSLAGT	60	
QY	61	YINYLGPPFNPBPDPNPPRLGAETLPRATVDLEVWRSNDKLRLTQNYEASHLLCYVRLG	120	
Db	61	YINYLGPPFNPBPDPNPPRLGAETLPRATVDLEVWRSNDKLRLTQNYEASHLLCYVRLG	120	
QY	121	NRQAATAELRSLAHFCTSLGGLGSGTAGWMAALGYPLPQPLPGTPTPTWPGPAISDFLQ	180	
Db	121	NRQAATAELRSLAHFCTSLGGLGSGTAGWMAALGYPLPQPLPGTPTPTWPGPAISDFLQ	180	
QY	181	KMDDFWLLKELQTWLWRSKDFNRLKKKQMPAAAAVTLHLGAHGF	225	
Db	181	KMDDFWLLKELQTWLWRSKDFNRLKKKQMPAAAAVTLHLGAHGF	225	

RESULT 7  
AAM25831  
ID AAM25831 standard; Protein; 253 AA.

AAM25831;

16-OCT-2001 (first entry)

Human protein sequence SEQ ID NO:1346.

Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia; antiaggregant; haemostatic; vulnery; antitumor; osteopathic; eczema; dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac synphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopaenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder.

Homo sapiens.

X

WO200153455-A2.

26-JUL-2001.

22-DEC-2000; 2000WO-US35017.

23-DEC-1999; 99US-0471275.

21-JAN-2000; 2000US-0488725.  
25-APR-2000; 2000US-0552317.

(HYSE-) HYSEO INC.

Tang YT, Lin C.

WPT: 2001-457603/19

N-PSDB; AAH99772.

Isolated human po...  
treatment and disc...

Claim 20; Page 278; 1217pp; English.

AAH99166 to AAH99904 encode the human proteins given in AAM25225 to AAM25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antiinflammatory; antirheumatic; antianthratic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; antiaggregant; haemostatic; vulnery; antitumor; osteopathic; dermatologic; antiangiogenic; antidiabetic; cystostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmunity, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders.

Sequence 253 AA;

ry Match	100.0%;	Score 1226;	DB 22;	Length 253;
Local Similarity	100.0%;	Pred. No. 1.6e-118;		

1 MDLRAGDSWGMLACLTVLWHLPAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLRSLAGT 60

29 MDLRAGDSWGLACLTVLWHLPAVPALNRTGDPGPGPSIQKTYDYLTRYLEHQLRSLAGT 88

61 YLN YLG PPF NE P DFN PRL GAETL PRATVDLE VWRSLNDKLR LTONYEAYSHLLCYLRGL 120

89 YLN YLGPPFNEPDFNPRLGAETLPRA TVDLEVWRS LNDKLR LRTQNYEAYSHLLCYLRGL 148

121 NRQAA TAE LRR SLA HCT SLQ LLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSD FLO 180

149 NRQAAATAE LRRSLAHFCTSLQGLLSIAGVMAALGYPLQPLPGTEPTWTPGPAHSDFLO 208

181 KMDDFWLLKEIQTLWRSKDFNRLKKKMQPPAAAVTLHLGAHGF 225

209 KMDDFWLLKELQTLWRS AKDFNRLKKKMPPAAA VTLHLGAHGF 253

88

28

AEV0620 standard; Protein; 223 AA.

AE00828;

DT	02-JUL-2001	(first entry)	ID	AB11896	standard; peptide; 321 AA.
XX			XX	AB11896;	
DE			AC		
XX			XX		
XX			DT		
KW	Human; biologically active complex; haemopoietin receptor; NR6;		XX	11-JAN-2002	(first entry)
KW	cardiotrophin-like cytokine; CLC; therapy; prophylaxis; proliferation;		DE		
KW	differentiation; cell survival; neurotrophic activity.		XX		
XX			XX		Human cardiotrophin-like cytokine homologue, SEQ ID NO:2266.
OS	Homo sapiens.		XX		
XX			KW		Human; cytokine; cell proliferation; cell differentiation; growth factor;
FT	Key	Location/Qualifiers	KW		haematopoiesis regulation; tissue growth; immunomodulator; activin;
FT	Peptide	1..27	KW		inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
FT	Protein	28..223	KW		proliferation; metastasis; cancer; tumour; haematopoietic disorder;
FT		/label= "Signal peptide"	KW		myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
FT		/label= "Human mature CLC protein"	KW		chronic inflammatory condition; proliferative retinopathy;
FT		/note= "Cardiotrophin-like cytokine"	KW		atherosclerosis; coronary heart disease; arterial ischaemia;
XX			KW		bone disorder; osteoporosis; vascular growth disorder;
XX	WO200127157-A1.		KW		tissue regeneration; wound healing; infection; immune disorder;
PN			KW		cell culture; drug screening; gene therapy; antiinflammatory;
XX			KW		antialthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
PD	19-APR-2001.		KW		cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
XX			KW		antifungal; vulnery; antiulcer.
PF	06-OCT-2000; 2000WO-AU01216.		XX		
XX			OS		Homo sapiens.
XX	08-OCT-1999; 99AU-0003327.		XX		
PR	12-MAY-2000; 2000AU-0007489.		XX		WO200157188-A2.
XX			XX		
XX	(AMRA-) AMRAD OPERATIONS PTY LTD.		PD		09-AUG-2001.
PA			XX		
XX			XX		05-FEB-2001; 2001WO-US03800.
XX	Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;		PF		
PI	Nakata Y, Hasegawa M;		XX		03-FEB-2000; 2000US-0496914.
XX			XX		27-APR-2000; 2000US-0560875.
DR	WPI; 2001-281978/29.		PR		
DR	N-PSDB; AAD04201.		XX		
XX			XX		(HYSE-) HYSEQ INC.
XX			XX		
PT	New biologically active complex comprising NR6 and		XX		
PT	cardiotrophin-like-cytokine, for facilitating proliferation,		PI		Tang YT, Liu C, Drmanac RT;
PT	differentiation and/or survival of a cell -		XX		
XX			XX		WPI; 2001-457740/49.
XX			DR		N-PSDB; ABA09140.
PS	Claim 32; Page 114-115; 123pp; English.		XX		
XX			XX		Human proteins and DNA encoding sequences useful for preventing,
CC	The present invention relates to a biologically active complex comprising		PT		treating or ameliorating a medical condition in a mammalian subject
CC	a haemopoietin receptor, NR6 and cardiotrophin-like cytokine (CLC).		PT		e.g. arthritis and cancer -
CC	The complex is useful in the manufacture of a medicament for the		XX		
CC	treatment and/or prophylaxis of a subject, as it is involved in		PS		Claim 20; Page 273; 1963pp; English.
CC	facilitating proliferation, differentiation and/or survival of a cell.		XX		
CC	The complex or its components have neurotrophic activity. The present		CC		Sequences AB10981-AB112330 represent 1350 novel human polypeptides, and
CC	sequence is human cardiotrophin-like cytokine (CLC) protein.		CC		sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
XX			CC		invention also relates to vectors and recombinant host cells comprising a
XX			CC		antibodies against the polypeptides, methods of producing the novel polypeptides,
XX			CC		nucleotide of the invention, methods of identifying compounds which
XX			CC		bind to polypeptides of the invention. Although novel, many of the
XX			CC		polypeptides of the invention have homology to known proteins, thereby
XX			CC		giving an insight into their probable biological activities, and hence
XX			CC		potential therapeutic applications. The polypeptides of the invention may
XX			CC		have various activities, including cytokine, stem cell growth factor activity;
XX			CC		differentiation activities; tissue growth activity;
XX			CC		haematopoiesis regulatory activity; activin- or inhibin-related activities;
XX			CC		immunomodulatory activity; activin- or inhibin-related activities; or may be
XX			CC		chemotactic or chemokinetic activities; haemostatic, thrombotic or
XX			CC		thrombolytic activities; receptor or ligand activities; or may be
XX			CC		involved in oncogenesis, cancer cell proliferation or metastasis.
XX			CC		Depending on their biological activities, polypeptides and nucleotides of
XX			CC		the invention are useful for preventing, treating or ameliorating medical
XX			CC		conditions, e.g., by protein or gene therapy. Such conditions include
XX			CC		cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
XX			CC		disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
XX			CC		proliferative retinopathy, atherosclerosis, coronary heart disease,
XX			CC		arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
XX			CC		vascular growth. Polypeptides involved with tissue regeneration and
XX			CC		repair (or nucleic acids encoding them) may be used to promote wound
XX			CC		healing (e.g., of burns, incisions and ulcers), while those with

immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

Sequence 321 AA;

Query March 98.2%; Score 1204; DB 22; Length 321;  
Best Local Similarity 99.1%; Pred. No. 4.2e-116;  
Matches 221; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LRAGDSWGLACLTVLWHLPAVPALNRTGDPGPGPSIQKTYDTRYLHQLRSLAGTYL 62  
DB 99 LPTGDSWGLACLTVLWHLPAVPALNRTGDPGPGPSIQKTYDTRYLHQLRSLAGTYL 158  
QY 63 NYLGPPNEPDPNPRGLCAETLPRATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 122  
DB 159 NYLGPPNEPDPNPRGLCAETLPRATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 218  
QY 123 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTTPGPAHSDFLQKM 182  
DB 219 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTTPGPAHSDFLQKM 278  
QY 183 DDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225  
DB 279 DDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 321

RESULT 10  
AAM79399  
ID AAM79399 standard; Protein; 321 AA.  
XX AC AAM79399;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human protein SEQ ID NO 3045.  
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX OS Homo sapiens.  
XX PN WO200157190-A2.  
XX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US04098.  
XX PR 03-FEB-2000; 2000US-0496914.  
XX PR 27-APR-2000; 2000US-0560875.  
XX PR 20-JUN-2000; 2000US-0598075.  
XX PR 19-JUL-2000; 2000US-0620325.  
XX PR 01-SEP-2000; 2000US-0654936.  
XX PR 15-SEP-2000; 2000US-0663561.  
XX PR 20-OCT-2000; 2000US-0693325.  
XX PR 30-NOV-2000; 2000US-0728422.  
XX PA (HYSB-) HYSEQ INC.  
XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZH;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX WPI; 2001-476283/51.

DR N-PSDB; AAK52532.

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX Claim 20; Page 237; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

XX Sequence 321 AA;

Query Match 98.2%; Score 1204; DB 22; Length 321;  
Best Local Similarity 99.1%; Pred. No. 4.2e-116;  
Matches 221; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LRAGDSWGLACLTVLWHLPAVPALNRTGDPGPGPSIQKTYDTRYLHQLRSLAGTYL 62  
DB 99 LPTGDSWGLACLTVLWHLPAVPALNRTGDPGPGPSIQKTYDTRYLHQLRSLAGTYL 158  
QY 63 NYLGPPNEPDPNPRGLCAETLPRATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 122  
DB 159 NYLGPPNEPDPNPRGLCAETLPRATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 218  
QY 123 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTTPGPAHSDFLQKM 182  
DB 219 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTTPGPAHSDFLQKM 278  
QY 183 DDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225  
DB 279 DDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 321

RESULT 11  
AAM78415  
ID AAM78415 standard; Protein; 260 AA.

XX AC AAM78415;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 1077.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

Sun Feb 2 08:31:46 2003

PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
 XX WPI; 2001-476283/51.  
 DR N-PSDB; AAK51548.  
 XX Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX Claim 20; Page 3306; 6221pp; English.  
 PS The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAW80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 XX Sequence 260 AA;  
 PS Query Match 97.6%; Score 1197; DB 22; Length 260;  
 XX Best Local Similarity 98.7%; Pred. No. 1.7e-115;  
 XX Matches 220; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 LRAGDSWGMACLTVMHLPAPVAPALNRTGDPGPGSIQKTYDLYLEHQLRSLAGTYL 62  
 Db 38 LPTGDSWGMACLTVMHLPAPVAPALNRTGDPGPGSIQKTYDLYLEHQLRSLAGTYL 97  
 QY 63 NYLGPPNEPDPNPRGLAETLPATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 122  
 Db 98 NYLGPPNEPDPNPRGLAETLPATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 157  
 QY 123 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQRM 182  
 Db 158 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQRM 217  
 QY 183 DDFWLLKELQTLWLRSAKDFNLRKKMQPPAAAVTLHLGAHGF 225  
 Db 218 DDFWLLKELQTLWLRSAKDFNLRKKMQPPAAAVTLHLGAHGF 260  
 RESULT 12  
 AAW29716 ID AAW29716 standard; Protein; 225 AA.  
 XX AAW29716;  
 AC AAW29716;  
 XX 09-NOV-1998 (first entry)  
 XX Mouse neurotrophic factor NNT-1.  
 DE NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;  
 XX Alzheimer's disease; Parkinson's disease; Huntington's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
 KW peripheral neuropathy; dystrophy; neural retina degeneration;  
 KW common variable immunodeficiency; CVID; selective IGA deficiency;  
 KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;  
 KW therapy.

XX Mus sp.  
 OS Key Location/Qualifiers  
 XX Peptide 1..27  
 FT Protein 28..225  
 FT /label= Sig\_peptide  
 FT /label= Mat\_protein  
 XX WO9833922-A1.  
 XX 06-AUG-1998.  
 XX 02-FEB-1998; 98WO-US02363.  
 XX 30-JAN-1998; 98US-0016534.  
 PR 03-FEB-1997; 97US-0792019.  
 XX (AMGE-) AMGEN INC.  
 XX Chang M, Elliot GS, Sarmiento U, Senaldi G;  
 WPI; 1998-437475/37.  
 N-PSDB; AAW47512.  
 XX Newly isolated nucleic acid encoding human or murine neurotrophic  
 PT factor NNT-1 - useful for treatment of neurological and  
 PT immunological diseases or inflammation, also as vaccine adjuvant  
 XX Claim 13; Fig 5; 120pp; English.  
 XX This is the amino acid sequence of a murine neurotrophic factor,  
 CC designated NNT-1, that is a growth factor for neurons and for B or  
 CC T cells. It was deduced from isolated NNT-1 cDNA (see AAW47512).  
 CC Human NNT-1 (see AAW29715) is also provided. Vectors and host cells  
 CC for use in the production of human murine recombinant NNT-1  
 CC polypeptides. These are used to treat: (i) neurological or  
 CC immunological diseases, specifically Alzheimer's, Parkinson's  
 CC or Huntington's diseases, amyotrophic lateral sclerosis,  
 CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and  
 CC degeneration of the neural retina, or conditions characterised by T  
 CC or B cell defects, e.g. common variable immunodeficiency (CVID),  
 CC selective IGA deficiency, hypogammaglobulinaemia and X-linked  
 CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)  
 CC inflammation. NNT-1 is also able to boost immunoreactivity and  
 CC antibody production following vaccination, and, since it inhibits  
 CC tumour necrosis factor production, it may also be useful for  
 CC treating sepsis. In addition, cells that have been engineered to  
 CC express NNT-1 can be implanted, or nucleic acids are delivered in  
 CC gene therapy vectors.  
 XX Sequence 225 AA;  
 PS Query Match 97.3%; Score 1193; DB 19; Length 225;  
 XX Best Local Similarity 96.9%; Pred. No. 3.6e-115;  
 XX Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MDLRAGDSWGMACLTVMHLPAPVAPALNRTGDPGPGSIQKTYDLYLEHQLRSLAGT 60  
 Db 1 MDLRAGDSWGMACLTVMHLPAPVAPALNRTGDPGPGSIQKTYDLYLEHQLRSLAGT 60  
 QY 61 YLYNGPPNEPDPNPRGLAETLPATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGL 120  
 Db 61 YLYNGPPNEPDPNPRGLAETLPATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGL 120  
 QY 121 NROQATAEELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180  
 Db 121 NROQATAEELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180  
 QY 181 KMDDFWLLKELQTLWLRSAKDFNLRKKMQPPAAAVTLHLGAHGF 225  
 Db 181 KMDDFWLLKELQTLWLRSAKDFNLRKKMQPPAAAVTLHLGAHGF 225



```
RESULT 13
AAW56142
ID AAW56142 standard; Protein; 225 AA.
XX
AC AAW56142;
XX
DT 13-JUL-1998 (first entry)
XX
DE Amino acid sequence of murine neurotrophic factor NNT-1.
XX
KW Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
KW treatment; neurological disease; degeneration; Parkinson's disease;
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT Peptide 1..27
FT Protein /note= "signal peptide"
FT Protein 28..225
FT Protein /note= "mature peptide"
XX
XX
PN US5741772-A.
XX
PD 21-APR-1998.
XX
PF 03-FEB-1997; 97US-0792019.
XX
PR 03-FEB-1997; 97US-0792019.
XX
PA (AMGE-) AMGEN INC.
XX
PI Chang M;
XX
DR WPI; 1998-260526/23.
DR N-PSDB; AAW22654.
XX
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
XX useful for stimulating growth of motor and sympathetic neurons
XX
XX Claim 2; Fig 5; 41pp; English.
XX
CC The present sequence represents a murine neurotrophic factor, designated
CC NNT-1, which is capable of stimulating growth of motor or sympathetic
CC neurons. The NNT-1 protein is useful in the treatment of neurological
CC diseases characterised by the degeneration and death of particular
CC classes of neurons. These diseases specifically include Parkinson's
CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
CC stroke and various degenerative disorders affecting vision.
XX
SQ Sequence 225 AA;
Query Match 97.3%; Score 1193; DB 19; Length 225;
Best Local Similarity 96.9%; Pred. No. 3.6e-115;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 MDLRAGDSWGMCLACTVWLHPALNRTGDPGPGSIQKTYDLTRYLHQRLSLAGT 60
DB 1 MDLRAGDSWGMCLACTVWLHPALNRTGDPGPGSIQKTYDLTRYLHQRLSLAGT 60
QY 61 YLNYLGPPFPNPPRLGATLPRATVDLEVWRSNDKRLRLTQNYEAYSHLLCYLRGL 120
DB 61 YLNYLGPPFPNPPRLGATLPRATVDLEVWRSNDKRLRLTQNYEAYSHLLCYLRGL 120
QY 121 NRQAATLRLRSIAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180
DB 121 NRQAATLRLRSIAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180
QY 181 KMDDFWLLKELQTLWRSADFNRLKKMQPPAAAVTLHLGAHGF 225
DB 181 KMDDFWLLKELQTLWRSADFNRLKKMQPPAAAVTLHLGAHGF 225
```

```
RESULT 14
AAW87814
ID AAW87814 standard; Protein; 225 AA.
XX
AC AAW87814;
XX
DT 24-AUG-2000 (first entry)
XX
DE Murine NNT-1 protein.
XX
KW NNT-1; neurotrophic factor; neurotropic; neuroprotective; treatment;
KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; murine;
KW Huntington's disease; peripheral neuropathy; neural retina degeneration;
KW retinopathy; immune disorder; hematopoietic disorder.
XX
OS Mus sp.
XX
XX US6054294-A.
XX
PD 25-APR-2000.
XX
PF 12-DEC-1997; 97US-0988819.
XX
PR 03-FEB-1997; 97US-0792019.
XX
PA (AMGE-) AMGEN INC.
XX
PI Chang M;
XX
DR WPI; 2000-338492/29.
DR N-PSDB; AAA39483.
XX
XX New nucleic acids encoding neurotrophic factors useful for stimulating
XX growth of motor or sympathetic neurons for treating neuron cell damage
XX
XX Claim 2b; Fig 5; 42pp; English.
XX
CC This invention describes a novel nucleic acid molecule (I) encoding a
CC novel neurotrophic factor (NNT-1) (II) which has neurotropic,
CC neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and
CC ophthalmological activity. (I) is useful for producing NNT-1
CC polypeptides which are useful for treating patients in whom various
CC cells of the central, autonomic, or peripheral nervous system have
CC degenerated and/or have been damaged by congenital disease, trauma,
CC mechanical damage, surgery, stroke, ischemia, infection, metabolic
CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1
CC proteins are used to treat diseases like Alzheimer's, Parkinson's,
CC amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's
CC disease, peripheral neuropathy induced by diabetes or other metabolic
CC disorders, and/or dystrophies or degeneration of the neural retina such
CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of
CC night blindness, progressive cone-rod degeneration, immune disorders and
CC hematopoietic disorders. (I) is effective in treating neurological
CC conditions and promotes neuron regeneration. Neural functions are
CC effectively restored in patients suffering from various neurological
CC disorders. This sequence represents the murine NNT-1 protein described in
CC the method of the invention.
XX
SQ Sequence 225 AA;
Query Match 97.3%; Score 1193; DB 21; Length 225;
Best Local Similarity 96.9%; Pred. No. 3.6e-115;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 MDLRAGDSWGMCLACTVWLHPALNRTGDPGPGSIQKTYDLTRYLHQRLSLAGT 60
DB 1 MDLRAGDSWGMCLACTVWLHPALNRTGDPGPGSIQKTYDLTRYLHQRLSLAGT 60
QY 61 YLNYLGPPFPNPPRLGATLPRATVDLEVWRSNDKRLRLTQNYEAYSHLLCYLRGL 120
```

CC immune system including rheumatoid arthritis, psoriatic arthritis,  
CC inflammatory arthritis, osteoarthritis, inflammatory joint disease,  
CC autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory  
CC bowel disease, transplant rejection, and graft versus host disease, and  
CC reproductive diseases and disorders including infertility, miscarriage,  
CC preterm labour and delivery, and endometriosis. The present sequence  
CC represents Mouse NNT1.  
XX  
SQ Sequence 225 AA;  
Query Match 97.3%; Score 1193; DB 23; Length 225;  
Best Local Similarity 96.9%; Pred. No. 3.6e-115;  
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MDLRAGDSWGMGLACTVLMHLPVPAVPAALNRTGDPGPGPSIQKTYDLYLTRYLHQLSLACT 60  
DB 1 MDLRAGDSWGMGLACTVLMHLPVPAVPAALNRTGDPGPGPSIQKTYDLYLTRYLHQLSLACT 60  
QY 61 YLNYLGPPFNEPDPNPRLGAETLPRATVLDLVWESLNDKRLTQNYEAYSHLLCYLRGL 120  
DB 61 YLNYLGPPFNEPDPNPRLGAETLPRATVLDLVWESLNDKRLTQNYEAYSHLLCYLRGL 120  
QY 121 NRQATAELRESLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTPGPAHSDFLQ 180  
DB 121 NRQATAELRESLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTPGPAHSDFLQ 180  
QY 181 KMDDFWLLKEQLTQWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225  
DB 181 KMDDFWLLKEQLTQWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225

Search completed: January 27, 2003, 15:41:34  
Job time : 35 secs

DB 61 YLNYLGPPFNEPDPNPRLGAETLPRATVLDLVWESLNDKRLTQNYEAYSHLLCYLRGL 120  
QY 121 NRQATAELRESLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTPGPAHSDFLQ 180  
DB 121 NRQATAELRESLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTPGPAHSDFLQ 180  
QY 181 KMDDFWLLKEQLTQWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225  
DB 181 KMDDFWLLKEQLTQWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225  
RESULT 15  
ID AAU78177 standard; Protein; 225 AA.  
XX  
AC AAU78177;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Mouse novel neurotrophic factor NNT1.  
XX  
KW Mouse; NNT1; neurotrophic factor; IgE-related disease;  
KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;  
KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;  
KW vascular restenosis; rheumatoid arthritis; psoriatic arthritis;  
KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;  
KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;  
KW inflammatory bowel disease; transplant rejection; reproductive disorder;  
KW graft versus host disease; infertility; miscarriage; preterm labour.  
XX  
OS Mus sp.  
XX  
XX WO200215977-A2.  
XX  
XX 28-FEB-2002.  
XX  
XX 17-AUG-2001; 2001WO-US25906.  
XX  
XX 19-AUG-2000; 2000US-226436P.  
XX  
XX 16-AUG-2001; 2001US-0931704.  
XX  
XX (AMGE-) AMGEN INC.  
XX  
XX Senaldi G;  
XX  
XX WPI; 2002-280867/32.  
XX  
XX N-PSDB; ABK11649.  
XX  
XX Treating Immunoglobulin E-related disease, modulating IgE levels in a  
XX patient, preventing IgE-related disease and treating allergic diseases,  
XX PT involves administering NNT-1 inhibitor to a patient -  
XX  
XX Claim 2; Fig 5; 63pp; English.

CC The invention relates to treating Immunoglobulin E (IgE)-related disease,  
CC modulating IgE levels in a patient, preventing an IgE-related disease,  
CC and treating allergic diseases, comprising administering a  
CC therapeutically effective amount of novel neurotrophic factor (NNT)-1  
CC inhibitor to a patient. Also included are a method of diagnosing an  
CC IgE-related disease or susceptibility to an IgE-related disease, by  
CC determining the presence or amount of expression of an NNT1 polypeptide  
CC encoded by a NNT1 nucleotide sequence, its fragment or naturally  
CC occurring variant, and diagnosing an IgE-related disease or  
CC susceptibility of an IgE-related disease based on the presence or amount  
CC of expression of the polypeptide and a pharmaceutical composition for use  
CC in treating IgE-related disease, comprising the NNT1 inhibitor.  
CC The NNT1 inhibitor is useful for preventing and treating IgE-related  
CC disease, modulating IgE levels, and treating allergic diseases e.g.  
CC Type I allergic disease, allergic rhinitis, eczema, dermatitis,  
CC pollinosis, asthma, immune diseases and disorders, diseases involving  
CC abnormal cell proliferation including cancer, arteriosclerosis and  
CC vascular restenosis, diseases and conditions relating to dysfunction of

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:42:42 ; Search time 9 Seconds

(without alignments)  
504.464 Million cell updates/sec

Title: US-09-931-704-2

Perfect score: 1226

Sequence: 1 MDLRAGDSWGLACLTCLVLM.....KKKMQPPAAAVTLHLGAHGF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pap.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pap.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pap.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pap.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pap.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1226	100.0	225	10	US-09-931-704-2
2	1193	97.3	225	10	US-09-931-704-5
3	885	72.2	154	10	US-09-864-761-40014
4	118.5	9.7	203	10	US-09-896-856-3
5	96.5	7.9	243	9	US-10-000-776-6
6	96.5	7.9	243	9	US-09-791-497-8
7	96.5	7.9	243	10	US-09-810-052-5
8	91.5	7.5	201	10	US-09-901-540-3
9	91.5	7.5	201	10	US-09-896-856-8
10	91.5	7.5	201	10	US-09-901-257-3
11	91	7.4	195	10	US-09-770-361-5
12	91	7.4	242	9	US-10-000-776-2
13	91	7.4	242	9	US-09-791-497-2
14	90	7.3	232	10	US-09-810-052-2
15	83	6.8	625	10	US-09-771-161A-242
16	83	6.8	625	10	US-09-771-161A-243
17	82	6.7	200	10	US-09-770-361-8
18	81.5	6.6	218	10	US-09-893-737-28
19	81	6.6	200	10	US-09-770-361-6

Sequence 4, Appli  
Sequence 12, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 6, Appli  
Sequence 10, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 10, Appli  
Sequence 16, Appli  
Sequence 16, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 8, Appli  
Sequence 10, Appli  
Sequence 60, Appli  
Sequence 2, Appli  
Sequence 1344, Ap  
Sequence 3, Appli  
Sequence 5603, Ap  
Sequence 12181, A  
Sequence 12995, A  
Sequence 13148, A

1399 9 US-09-388-221-4  
1424 9 US-09-388-221-12  
1429 10 US-09-996-617-2  
1429 10 US-09-931-071-2  
1443 9 US-09-388-221-6  
1443 9 US-09-388-221-10  
1454 9 US-09-388-221-2  
1473 9 US-09-388-221-2  
348 10 US-09-730-617-2  
418 10 US-09-946-807-3  
418 10 US-09-795-668-3  
418 10 US-09-795-686-3  
200 10 US-09-770-361-10  
379 9 US-09-860-846-16  
379 10 US-09-861-289-16  
3782 9 US-09-860-846-4  
3782 10 US-09-861-289-4  
234 9 US-10-000-776-8  
234 9 US-09-791-497-10  
234 9 US-10-078-929-60  
439 12 US-10-124-429-2  
416 9 US-09-764-864-1344  
451 10 US-09-770-361-3  
199 10 US-09-815-242-5603  
917 10 US-09-815-242-12181  
920 10 US-09-815-242-12995  
920 10 US-09-815-242-13148

#### ALIGNMENTS

##### RESULT 1

US-09-931-704-2  
; Sequence 2, Application US/09931704  
; Patent No. US20020041873A1  
; GENERAL INFORMATION:  
; APPLICANT: Senaldi, Giorgio  
; TITLE OF INVENTION: Methods and Compositions for Treating IGE-Related Disease Using  
; TITLE OF INVENTION: Inhibitors  
; FILE REFERENCE: A-695  
; CURRENT APPLICATION NUMBER: US/09/931,704  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: US 60/226,436  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-931-704-2

Query Match 100.0%; Score 1226; DB 10; Length 225;  
Best Local Similarity 100.0%; Pred. No. 6.5e-113;  
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGLACLTCLVLMHLPAPVPAALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60  
DB 1 MDLRAGDSWGLACLTCLVLMHLPAPVPAALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60

QY 61 YLYNLGPPFPNPPRLGAETLPRATVLEWVRSNDKRLKRLTQNYEAYSHLLCYLRGL 120  
DB 61 YLYNLGPPFPNPPRLGAETLPRATVLEWVRSNDKRLKRLTQNYEAYSHLLCYLRGL 120

QY 121 NRQATAELRSLAHFTCSLQGLLSTAGVMAALGYLPQPLPCTETPTWTPGAHSDFLQ 180  
DB 121 NRQATAELRSLAHFTCSLQGLLSTAGVMAALGYLPQPLPCTETPTWTPGAHSDFLQ 180

QY 181 KMDDFWLLKELQTLWRSKDFNRLKKKMQPPAAAVTLHLGAHGF 225  
DB 181 KMDDFWLLKELQTLWRSKDFNRLKKKMQPPAAAVTLHLGAHGF 225

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40014
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: HIT: AI752561.1, EVALUUE 3.00e-66
; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUUE 8.00e-03
US-09-864-761-40014

Query Match 72.2%; Score 885; DB 10; Length 164;
Best Local Similarity 99.4%; Pred. No. 1.2e-79;
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0

Qy 62 LNYLGGPFNFDPNPRLGAETLPRTATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGLN 121
Db 1 LNYLGGPFNFDPNPRLGAETLPRTATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGLN 60

Qy 122 RQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPDLPTETPTWPGPAHSDFLOK 181
Db 61 RQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPDLPTETPTWPGPAHSDFLOK 120

Qy 182 MDDFWLLKELQTLWRSKADFNELKKMQPPAAAVTLHLGAHGF 225
Db 121 MDDFWLLKELQTLWRSKADFNELKKMQPPAAAVTLHLGAHGF 164

RESULT 4
US-09-896-856-3
; Sequence 3, Application US/09896856
; Patent No. US20020137189A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; Chien, Kenneth
; King, Kathleen
; Pennica, Diane
; Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California

```



```
QY 47 TRYLEHQLRSLAGTYLN-----YLGPPFNEPDPNPRRLGAETLPRAATVDLEVMWSLN 98
Db 54 ARKLSEVRGQHRFAESHLPGVNLVLP-----LG-EQLPDVSLTFQAWRLS 101
QY 99 DKRLTONYEAYSHLLCYLRGLNRQAATAELRR-SLAHFCTSLQGLLSIAGVMAALGYP 157
Db 102 DPERLCFISTTLQPPHAPLGGLTQGRWTNMMERQMLWAMRLDLRDLQRHLRFQVLAAGFN 161
QY 158 LPQP-----LPQTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLSRAKDF 202
Db 162 LPPEEEEEEEEEERKGLLPGALGALQGPQVSWPQLLSTYRLLSLHSLVLSRAVREL 221
QY 203 NRLKK 207
Db 222 LLLSK 226

RESULT 7
US-09-810-052-5
; Sequence 5, Application US/09810052
; Patent No. US20020009775A1
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Preenell, Scott R.
; TITLE OF INVENTION: HELICAL PROTEIN ZALPHA51
; FILE REFERENCE: 00-24
; CURRENT APPLICATION NUMBER: US/09/810,052
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/190,410
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/199,443
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-810-052-5

Query Match 7.9%; Score 96.5; DB 10; Length 243;
Best Local Similarity 23.7%; Pred. No. 0.04;
Matches 58; Conservative 26; Mismatches 104; Indels 57; Gaps 9;

QY 1 MDLRAGD-SWGMACLCITVL-----WHLPVAPALNRTGDPGPGPSIQK-----TYDL 46
Db 1 MGQTAGDLGWRLSLLLLPLLVQAGVWGFRPPG-----RPQLSLQELRREFTVLSHL 53
QY 47 TRYLEHQLRSLAGTYLN-----YLGPPFNEPDPNPRRLGAETLPRAATVDLEVMWSLN 98
Db 54 ARKLSEVRGQHRFAESHLPGVNLVLP-----LG-EQLPDVSLTFQAWRLS 101
QY 99 DKRLTONYEAYSHLLCYLRGLNRQAATAELRR-SLAHFCTSLQGLLSIAGVMAALGYP 157
Db 102 DPERLCFISTTLQPPHAPLGGLTQGRWTNMMERQMLWAMRLDLRDLQRHLRFQVLAAGFN 161
QY 158 LPQP-----LPQTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLSRAKDF 202
Db 162 LPPEEEEEEEEEERKGLLPGALGALQGPQVSWPQLLSTYRLLSLHSLVLSRAVREL 221
QY 203 NRLKK 207
Db 222 LLLSK 226

RESULT 8
US-09-901-540-3
; Sequence 3, Application US/09901540
; Patent No. US20020102622A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Goddard, Audrey
```

```
; APPLICANT: Lawrence, David A.
; APPLICANT: Pennica, Diane
; APPLICANT: Roy, Margaret A.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Cardiophin-1 Compositions and Methods for the
; FILE REFERENCE: P2533D2
; CURRENT APPLICATION NUMBER: US/09/901,540
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 09/648,258
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: US 09/033,114
; PRIOR FILING DATE: 1998-03-02
; PRIOR APPLICATION NUMBER: US 08/733,850
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: US 08/443,129
; PRIOR FILING DATE: 1995-05-17
; PRIOR APPLICATION NUMBER: US 08/286,304
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: US 08/233,609
; PRIOR FILING DATE: 1994-04-25
; PRIOR APPLICATION NUMBER: US 60/113,296
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Human
US-09-901-540-3

Query Match 7.5%; Score 91.5; DB 10; Length 201;
Best Local Similarity 25.9%; Pred. No. 0.096;
Matches 45; Conservative 24; Mismatches 92; Indels 13; Gaps 5;

QY 40 IQKTYDLTRYLEHQLRSLAGTYLNLYLGPPFNEPDPNPRRLGAETLPRAATVDLEVMWSLN 99
Db 27 IROTHSLAHLTKYAEQLQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 55
QY 100 KLRLTONYEAYSHLLCYLRGLNRQAATAELRR-SLAHFCTSLQGLLSIAGVMAAL 154
Db 86 RLRL--DAALALPPLLDVCRQAELNPRAPRLRLLEDAARQARALGAVALAAL 143
QY 155 GYPLPQLPGTEP---TWTPGPAHSDFLQKMDDFWLLKELQTLWLSRAKDFNRL 205
Db 144 G--AANRGPRAEPAPATAASATGVPKAVGLRVCGLYREWLRSRTGDLQQL 195

RESULT 9
US-09-896-856-8
; Sequence 8, Application US/09896856
; Patent No. US20020137189A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; Chien, Kenneth
; King, Kathleen
; Pennica, Diane
; Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/896,856
```

;; FILING DATE: 29-Jun-2001  
;; CLASSIFICATION DATA: <Unknown>  
;; PRIOR APPLICATION NUMBER: US 08/733,850  
;; FILING DATE: 18-OCT-1996  
;; APPLICATION NUMBER: US 08/471,112  
;; FILING DATE: 06-JUN-1995  
;; APPLICATION NUMBER: 08/233,609  
;; FILING DATE: 25-APR-1994  
;; APPLICATION NUMBER: 08/286304  
;; FILING DATE: 05-AUG-1994  
;; APPLICATION NUMBER: 08/443129  
;; FILING DATE: 17-MAY-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Conley, Deirdre L.  
;; REGISTRATION NUMBER: 36,487  
;; REFERENCE/DOCKET NUMBER: P0894PID2C1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415/225-2066  
;; TELEFAX: 415/952-9881  
;; TELEX: 910/371-7168  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 201 amino acids  
;; TYPE: Amino Acid  
;; TOPOLOGY: Linear  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-896-856-8

Query Match 7.5%; Score 91.5; DB 10; Length 201;  
Best Local Similarity 25.9%; Pred. No. 0.096; Indels 13; Gaps 5;  
Matches 45; Conservative 24; Mismatches 92;  
QY 40 IQKTVDLTRYLEHQRLSLAGTYLNYLGPPFNEPDPNPPRLGAETLPRAIVDLEWVRSND 99  
DB 27 IRQTHSLAHLTKYAEQLLEQVQLQGGDFGLPSFPPRLPVAGL-SAPAPSHAGLPVHE 85  
QY 100 KLRLTONYEAYSHLLCYLRGLNRQAA-----TAEIIRSLAHFCTSLQGLLSIAGVMAAL 154  
DB 86 RLRL--DAAALAAPLLDVCRCRQAEINPRAPRLRLLEDAAARQARALGAAVEALLAAL 143  
QY 155 GYPLPQPLPGTEP---TWTPGPAHSDFLQKMDDFWLLKELQTLWRSKDFNRL 205  
DB 144 G--AANRGPRAPPAATASASATGVFPKVLGLRVCGLYREWLSTRTEGDLGQL 195

RESULT 10  
US-09-901-257-3  
; Sequence 3, Application US/09901257  
; Patent No. US20020146707A1  
; GENERAL INFORMATION:  
; APPLICANT: Botstein, David  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Lawrence, David A.  
; APPLICANT: Pennica, Diane  
; APPLICANT: Roy, Margaret A.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: Cardiotrophin-1 Compositions and Methods for the  
; FILE OF INVENTION: Treatment of Tumor  
; FILE REFERENCE: P2533D3  
; CURRENT FILING DATE: US/09/901,257  
; PRIOR APPLICATION NUMBER: 09/648,252  
; PRIOR FILING DATE: 2000-08-25  
; PRIOR APPLICATION NUMBER: US 09/033,114  
; PRIOR FILING DATE: 1998-03-02  
; PRIOR APPLICATION NUMBER: US 08/733,850  
; PRIOR FILING DATE: 1996-10-18  
; PRIOR APPLICATION NUMBER: US 08/443,129  
; PRIOR FILING DATE: 1995-05-17  
; PRIOR APPLICATION NUMBER: US 08/286,304  
; PRIOR FILING DATE: 1994-08-05  
; PRIOR APPLICATION NUMBER: US 08/233,609

;; PRIOR FILING DATE: 1994-04-25  
;; PRIOR APPLICATION NUMBER: US 60/113,296  
;; PRIOR FILING DATE: 1998-12-22  
;; NUMBER OF SEQ ID NOS: 6  
;; SEQ ID NO 3  
;; LENGTH: 201  
;; TYPE: PRT  
;; ORGANISM: Human  
US-09-901-257-3

Query Match 7.5%; Score 91.5; DB 10; Length 201;  
Best Local Similarity 25.9%; Pred. No. 0.096; Indels 13; Gaps 5;  
Matches 45; Conservative 24; Mismatches 92;  
QY 40 IQKTVDLTRYLEHQRLSLAGTYLNYLGPPFNEPDPNPPRLGAETLPRAIVDLEWVRSND 99  
DB 27 IRQTHSLAHLTKYAEQLLEQVQLQGGDFGLPSFPPRLPVAGL-SAPAPSHAGLPVHE 85  
QY 100 KLRLTONYEAYSHLLCYLRGLNRQAA-----TAEIIRSLAHFCTSLQGLLSIAGVMAAL 154  
DB 86 RLRL--DAAALAAPLLDVCRCRQAEINPRAPRLRLLEDAAARQARALGAAVEALLAAL 143  
QY 155 GYPLPQPLPGTEP---TWTPGPAHSDFLQKMDDFWLLKELQTLWRSKDFNRL 205  
DB 144 G--AANRGPRAPPAATASASATGVFPKVLGLRVCGLYREWLSTRTEGDLGQL 195

RESULT 11  
US-09-770-361-5  
; Sequence 5, Application US/09770361  
; Patent No. US20020123462A1  
; GENERAL INFORMATION:  
; APPLICANT: Fandl, James  
; APPLICANT: Stahl, Neil  
; TITLE OF INVENTION: Modified Ciliary Neurotrophic Factor, Method Of Making  
; FILE OF INVENTION: And Methods Of Use Thereof  
; FILE REFERENCE: REG142  
; CURRENT APPLICATION NUMBER: US/09/770,361  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 09/031,693  
; PRIOR FILING DATE: 1998-02-27  
; PRIOR APPLICATION NUMBER: 08/308,736  
; PRIOR FILING DATE: 1994-09-19  
; PRIOR APPLICATION NUMBER: 07/959,284  
; PRIOR FILING DATE: 1992-10-09  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 195  
; TYPE: PRT  
; ORGANISM: CHICKEN  
US-09-770-361-5

Query Match 7.4%; Score 91; DB 10; Length 195;  
Best Local Similarity 27.3%; Pred. No. 0.1; Indels 32; Gaps 9;  
Matches 51; Conservative 21; Mismatches 83;  
QY 46 LTRYLEHQRLSLAGTYLNYLGPPFNEPDPNPPRLGAETLPRAIVDLEWVRSNDKRLRTO 105  
DB 23 LARKMSRSDVTDLLDIYVERQG-----LDASISVAADVGPPTAAV--ERWAEQTGTQRLLD 75  
QY 106 N---YEAYSHLLCYLRGLNRQAA-----ATAELRSLA-----HFCTSIQGLLSIAGVMA 152  
DB 76 NLAAYRAFRITLLAQMLEEQEQLLEDGTDRLGPAALAMLLQVSAPVYHLEEL-----ELE 130  
QY 153 ALGYPLPQPLPGTEPTWTPGAH--SDFLQKMDDFWLLKELQTLWRSKDFNRLKKMQP 211  
DB 131 SRGAPAE---GSEP---PAPPRLSLFEQKLRGLRVRLRELAQWAVRSVRDLRLQLSKHGP 184  
QY 212 PAAAVTL 218  
DB 185 SGAALGL 191

## RESULT 12

US-10-000-776-2

; Sequence 2, Application US/10000776

; Patent No. US20020164609A1

; GENERAL INFORMATION:

; APPLICANT: Timans, Jacqueline C.

; APPLICANT: Pflanz, Stefan K.-H.

; APPLICANT: Kastelein, Robert A.

; APPLICANT: Bazan, Jose F.

; APPLICANT: Rennick, Donna

; APPLICANT: de Waal Malefyt, Rene

; APPLICANT: Cheung, Jeanne

; TITLE OF INVENTION: MAMMALIAN CYTOKINES; RELATED REAGENTS

; FILE REFERENCE: DX01040K3

; CURRENT APPLICATION NUMBER: US/10/000,776

; CURRENT FILING DATE: 2001-11-30

; PRIOR APPLICATION NUMBER: 09/791,497

; PRIOR FILING DATE: 2001-02-22

; PRIOR APPLICATION NUMBER: 09/627,897

; PRIOR FILING DATE: 2000-07-27

; PRIOR APPLICATION NUMBER: 60/146,581

; PRIOR FILING DATE: 1999-07-30

; PRIOR APPLICATION NUMBER: 60/147,763

; PRIOR FILING DATE: 1999-08-06

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 2

; LENGTH: 242

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-000-776-2

Query Match 7.4%; Score 91; DB 9; Length 242;

Best Local Similarity 23.0%; Pred. No. 0.14;

Matches 56; Conservative 26; Mismatches 105; Indels 56; Gaps 8;

QY 2 DLKAGDSWGLACLCTVL-----WHLPAVPALNRTGDPGPGPSIQK-----TYDLTR 48  
DB 2 DLENNPKIGLSLLPLLLVQAGVWGFPFPPG-----RQLSLQELRRFTVSLHLAR 54  
QY 49 YLEHQLRSLAGTYLN-----YLGPPFNEPDPNPRPGAETLPRATVDLEWVRSNDK 100  
DB 55 KLLSEVRGQAHRAESHLPGVNLVLLP-----LG-EQLPDVSLTFQAWRRLSDP 102  
QY 101 LRLTQNYEAYSHLLCYLRGLNRQAATAELRR-SLAHFCTSLQGLGSIAGVMAALGYPLP 159  
DB 103 ERLCFISTTLQPPHAPLGGLGTQGRWTNMERMQLWAMRLDLRDLQRHLRFQVLAAGFNLP 162  
QY 160 QP-----LPGTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNR 204  
DB 163 EEEEEEEEEERKGLLPALGALQPAQVSWPQLLSTYRLHLSLELVLSRAVRELLL 222  
QY 205 LKK 207  
DB 223 LSK 225

## RESULT 13

US-09-791-497-2

; Sequence 2, Application US/09791497

; Publication No. US20030008343A1

; GENERAL INFORMATION:

; APPLICANT: Timans, Jacqueline C.

; APPLICANT: Kastelein, Robert A.

; APPLICANT: Bazan, J. Fernando

; APPLICANT: Pflanz, Stefan

; TITLE OF INVENTION: Mammalian Cytokines; Related Reagents

; FILE REFERENCE: DX01040K2

; CURRENT APPLICATION NUMBER: US/09/791,497

; CURRENT FILING DATE: 2001-02-22

; PRIOR APPLICATION NUMBER: 09/627,897

; PRIOR FILING DATE: 2000-07-27

; PRIOR APPLICATION NUMBER: 60/146,581

; PRIOR FILING DATE: 1999-07-30

; PRIOR APPLICATION NUMBER: 60/147,763

; PRIOR FILING DATE: 1999-08-06

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 2

; LENGTH: 242

; TYPE: PRT

; ORGANISM: primate; surmised Homo sapiens

US-09-791-497-2

Query Match 7.4%; Score 91; DB 9; Length 242;

Best Local Similarity 23.0%; Pred. No. 0.14;

Matches 56; Conservative 26; Mismatches 105; Indels 56; Gaps 8;

QY 2 DLKAGDSWGLACLCTVL-----WHLPAVPALNRTGDPGPGPSIQK-----TYDLTR 48  
DB 2 DLENNPKIGLSLLPLLLVQAGVWGFPFPPG-----RQLSLQELRRFTVSLHLAR 54  
QY 49 YLEHQLRSLAGTYLN-----YLGPPFNEPDPNPRPGAETLPRATVDLEWVRSNDK 100  
DB 55 KLLSEVRGQAHRAESHLPGVNLVLLP-----LG-EQLPDVSLTFQAWRRLSDP 102  
QY 101 LRLTQNYEAYSHLLCYLRGLNRQAATAELRR-SLAHFCTSLQGLGSIAGVMAALGYPLP 159  
DB 103 ERLCFISTTLQPPHAPLGGLGTQGRWTNMERMQLWAMRLDLRDLQRHLRFQVLAAGFNLP 162  
QY 160 QP-----LPGTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNR 204  
DB 163 EEEEEEEEEERKGLLPALGALQPAQVSWPQLLSTYRLHLSLELVLSRAVRELLL 222  
QY 205 LKK 207  
DB 223 LSK 225

## RESULT 14

US-09-810-052-2

; Sequence 2, Application US/09810052

; Patent No. US20020009775A1

; GENERAL INFORMATION:

; APPLICANT: Conklin, Darrell C.

; APPLICANT: Presnell, Scott R.

; TITLE OF INVENTION: HELICAL PROTEIN ZALPHA51

; FILE REFERENCE: 00-24

; CURRENT APPLICATION NUMBER: US/09/810,052

; CURRENT FILING DATE: 2001-03-16

; PRIOR APPLICATION NUMBER: 60/190,410

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: 60/199,443

; PRIOR FILING DATE: 2000-04-25

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 232

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-810-052-2

Query Match 7.3%; Score 90; DB 10; Length 232;

Best Local Similarity 23.2%; Pred. No. 0.16;

Matches 51; Conservative 25; Mismatches 94; Indels 50; Gaps 7;

QY 19 LWHLPALNRTGDPGPGPSIQK-----TYDLTRYLEHQLRSLAGTYLN-----63  
DB 15 VNGFPPPG-----RQLSLQELRRFTVSLHLARLKLSEVRGQAHRAESHLPGVNL 67  
QY 64 YLGPPFNEPDPNPRPGAETLPRATVDLEWVRSNDKRLTQNYEAYSHLLCYLRGLNRQ 123  
DB 68 YLLP-----LG-EQLPDVSLTFQAWRRLSDPERLCFISTTLQPPHAPLGGLGTQ 115  
QY 124 AATAELRR-SLAHFCTSLQGLGSIAGVMAALGYPLPQP-----LPCTEP 167



Db 116 GRWTMMERQMLWAMRLDRLDQLRHFRFOVLAAGFNLPPEEEEEEEERKGLLPGLG 175  
Qy 168 TWTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNRLKK 207  
Db 176 SALQSPAQVSWPQLLSTYRLHLSLEVLRSRAVRELLLSK 215

RESULT 15

US-09-771-161A-242  
; Sequence 242, Application US/09771161A  
; Patent No. US20020110811A1  
; GENERAL INFORMATION:  
; APPLICANT: LEVINE, et al.  
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES  
; FILE REFERENCE: 802620-2005.1  
; CURRENT APPLICATION NUMBER: US/09/771.161A  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 09/724,676  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 136776  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 135619  
; PRIOR FILING DATE: 2000-04-12  
; NUMBER OF SEQ ID NOS: 273  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 242  
; LENGTH: 625  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-771-161A-242

Query Match 6.8%; Score 83; DB 10; Length 625;  
Best Local Similarity 25.2%; Pred. No. 2.8;  
Matches 60; Conservative 26; Mismatches 78; Indels 74; Gaps 15;  
Qy 2 DLKAGDSGMLACLCTVLWHLPAV-----PALNRTGDPGPGPSIQKTYDLTRYLEHQLR 55  
Db 382 DIREGAPLGV-----HLFFVGYSYSCMALRDSEVPGPTP-----MELEAE 421  
Qy 56 SLAGTYLNYLGLPPFNEPDNPPRLGAE-----TLPRATVDLEV-WRSINDKLR---LTQN 106  
Db 422 QLLEPHVQ---APSLEPSVSQDETAEVAVPAAPAAEAETVIRELQEPLEEEVLTR- 477  
Qy 107 YEAYSHLLCYLRGNLQRAAT---AELR-RSLAHFCTSLQGLLGI--AGVMAALGYPLP 159  
Db 478 -QSLREMEAIRTDNQNFASQLREAEARNRDLAEHVRQLQERMELLQAEAGATAVTGVESP 536  
Qy 160 QPLFGTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAAAVT 217  
Db 537 R---ATDP-----PSH-----MAPRPLWASARWVG-----QAPCTAAT 567

Search completed: January 27, 2003, 15:49:54  
Job time : 10 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:22 ; Search time 13 Seconds  
(without alignments)  
509.243 Million cell updates/sec

Title: US-09-931-704-2

Perfect score: 1226

Sequence: 1 MDLRAGDSWGMCLCTVLM.....KKKQPPAAAVTLHGAGHF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1226	100.0	225	1	US-08-792-019B-2
2	1226	100.0	225	3	US-08-106-182-2
3	1226	100.0	225	3	US-08-988-819-2
4	1226	100.0	225	4	US-09-016-534-2
5	1193	97.3	225	1	US-08-792-019B-5
6	1193	97.3	225	3	US-08-988-819-5
7	1193	97.3	225	4	US-09-016-534-5
8	124.5	10.2	203	3	US-08-106-182-3
9	118.5	9.7	203	1	US-08-233-609-3
10	118.5	9.7	203	1	US-08-444-083-3
11	118.5	9.7	203	1	US-08-286-304-3
12	118.5	9.7	203	1	US-08-442-745-3
13	118.5	9.7	203	1	US-08-443-129-3
14	118.5	9.7	203	1	US-08-443-952-3
15	118.5	9.7	203	1	US-08-443-130-3
16	118.5	9.7	203	3	US-08-898-911-3
17	118.5	9.7	203	5	PCT-US95-04467-3
18	91.5	7.5	201	1	US-08-444-083-8
19	91.5	7.5	201	1	US-08-286-304-8
20	91.5	7.5	201	1	US-08-442-745-8
21	91.5	7.5	201	1	US-08-443-129-8
22	91.5	7.5	201	1	US-08-443-952-8
23	91.5	7.5	201	1	US-08-443-130-8
24	91.5	7.5	201	1	US-08-792-019B-11
25	91.5	7.5	201	3	US-09-106-182-4
26	91.5	7.5	201	3	US-08-988-819-11
27	91.5	7.5	201	3	US-08-898-911-8

```

28 91.5 7.5 201 4 US-09-016-534-11 Sequence 11, Appli
29 91.5 7.5 201 5 PCT-US95-04467-8 Sequence 8, Appli
30 91.5 7.5 1182 4 US-09-287-354-6 Sequence 6, Appli
31 91 7.4 195 1 US-07-959-284-5 Sequence 5, Appli
32 91 7.4 195 2 US-08-308-736A-5 Sequence 5, Appli
33 91 7.4 195 4 US-08-645-107A-5 Sequence 5, Appli
34 91 7.4 195 4 US-09-197-349-5 Sequence 2, Appli
35 91 7.4 195 5 PCT-US92-08258-2 Sequence 2, Appli
36 91 7.4 195 5 PCT-US93-09649A-5 Sequence 5, Appli
37 91 7.4 195 5 PCT-US93-09649-5 Sequence 5, Appli
38 87.5 7.1 1207 4 US-09-287-354-5 Sequence 5, Appli
39 85.5 7.0 560 2 US-08-756-317-11 Sequence 11, Appli
40 84 6.9 200 4 US-08-949-155-4 Sequence 4, Appli
41 84 6.9 200 4 US-09-819-964-4 Sequence 4, Appli
42 82.5 6.7 582 2 US-08-422-699A-9 Sequence 9, Appli
43 82.5 6.7 582 2 US-08-422-706B-9 Sequence 9, Appli
44 82.5 6.7 984 4 US-09-287-354-2 Sequence 2, Appli
45 82.5 6.7 1189 4 US-09-287-354-4 Sequence 4, Appli

```

## ALIGNMENTS

```

RESULT 1
US-08-792-019B-2
; Sequence 2, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUROTROPIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-792-019B-2

```

Query Match 100.0%; Score 1226; DB 1; Length 225;  
Best Local Similarity 100.0%; Pred. No. 3.7e-129;  
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MDLRAGDSWGMCLCTVLMHLPVAPALNRTGDPGSPSIQKTYDLTRYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGMCLCTVLMHLPVAPALNRTGDPGSPSIQKTYDLTRYLEHQLRSLAGT 60
QY 61 YLNYLGPPFPNPPRLGAETLPVATVLEWRSNDKRLRNTQYEAASHLLCYLRGL 120
Db 61 YLNYLGPPFPNPPRLGAETLPVATVLEWRSNDKRLRNTQYEAASHLLCYLRGL 120
QY 121 NRQATAEARRSLAHFCTSLQGLGSIAGVMAALGYPLQPLPGTEPTTPGPAHSDFLQ 180

```

Db 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTWTGPAHSDFLQ 180  
Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225  
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225

RESULT 2  
US-09-106-182-2  
; Sequence 2, Application US/09106182  
; Patent No. 6046035  
; GENERAL INFORMATION:  
; APPLICANT: Shi, Yangu  
; APPLICANT: Ruben, Steve  
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc  
; STREET: 9410 Key West Ave  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: US  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/106,182  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/051,053  
; FILING DATE: 30-JUN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF385  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-106-182-2

Query Match 100.0%; Score 1226; DB 3; Length 225;  
Best Local Similarity 100.0%; Pred. No. 3.7e-129;  
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMLACLTCTVWLHPVAPALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60  
Db 1 MDLRAGDSWGMLACLTCTVWLHPVAPALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60

Qy 61 YLYLGGPPFNEPDPNPPRLGAETLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRGL 120  
Db 61 YLYLGGPPFNEPDPNPPRLGAETLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRGL 120

Qy 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTWTGPAHSDFLQ 180  
Db 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTWTGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225  
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225

RESULT 3  
US-08-988-819-2

; Sequence 2, Application US/08988819  
; Patent No. 6054294  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/988,819  
; FILING DATE: 12-DEC-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442A  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-988-819-2

Query Match 100.0%; Score 1226; DB 3; Length 225;  
Best Local Similarity 100.0%; Pred. No. 3.7e-129;  
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMLACLTCTVWLHPVAPALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60  
Db 1 MDLRAGDSWGMLACLTCTVWLHPVAPALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60

Qy 61 YLYLGGPPFNEPDPNPPRLGAETLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRGL 120  
Db 61 YLYLGGPPFNEPDPNPPRLGAETLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRGL 120

Qy 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTWTGPAHSDFLQ 180  
Db 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTWTGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225  
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225

RESULT 4  
US-09-016-534-2  
; Sequence 2, Application US/09016534  
; Patent No. 6143874  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; APPLICANT: ELLIOTT, GARY S.  
; APPLICANT: SARMIENTO, ULLA  
; APPLICANT: SENALDI, GIORGIO  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER  
; CITY: THOUSAND OAKS

STATE: CA  
COUNTRY: USA  
ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,534  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/792,019  
FILING DATE: 03-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: COOK, ROBERT R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-442B  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-016-534-2

Query Match 100.0%; Score 1226; DB 4; Length 225;  
Best Local Similarity 100.0%; Pred. No. 3.7e-129;  
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACICTVLVHLPAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLSLAGT 60  
Db 1 MDLRAGDSWGLACICTVLVHLPAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLSLAGT 60  
Qy 61 YNLVIGPPFNEPDNPRLGAETLPRATVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120  
Db 61 YNLVIGPPFNEPDNPRLGAETLPRATVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120  
Qy 121 NRQAATAEIRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTEPTWTGPAHSDFLQ 180  
Db 121 NRQAATAEIRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTEPTWTGPAHSDFLQ 180  
Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKKQPPAAAATVTLHLGAHGF 225  
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKKQPPAAAATVTLHLGAHGF 225

RESULT 5  
US-08-792-019B-5  
Sequence 5, Application US/08792019B  
Patent No. 5741772  
GENERAL INFORMATION:  
APPLICANT: CHANG, MING-SHI  
TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMGEN INC.  
STREET: 1840 DEHAVILLAND DRIVE  
CITY: THOUSAND OAKS  
STATE: CA  
COUNTRY: USA  
ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/792,019B  
FILING DATE: 03-FEB-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:

NAME: COOK, ROBERT R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-442  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-792-019B-5

Query Match 97.3%; Score 1193; DB 1; Length 225;  
Best Local Similarity 96.9%; Pred. No. 1.8e-125;  
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACICTVLVHLPAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLSLAGT 60  
Db 1 MDLRAGDSWGLACICTVLVHLPAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLSLAGT 60  
Qy 61 YNLVIGPPFNEPDNPRLGAETLPRATVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120  
Db 61 YNLVIGPPFNEPDNPRLGAETLPRATVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120  
Qy 121 NRQAATAEIRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTEPTWTGPAHSDFLQ 180  
Db 121 NRQAATAEIRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTEPTWTGPAHSDFLQ 180  
Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKKQPPAAAATVTLHLGAHGF 225  
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKKQPPAAAATVTLHLGAHGF 225

RESULT 6  
US-08-988-819-5  
Sequence 5, Application US/08988819  
Patent No. 6054294  
GENERAL INFORMATION:  
APPLICANT: CHANG, MING-SHI  
TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMGEN INC.  
STREET: ONE AMGEN CENTER DRIVE  
CITY: THOUSAND OAKS  
STATE: CA  
COUNTRY: USA  
ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/988,819  
FILING DATE: 12-DEC-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/792,019  
FILING DATE: 03-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: COOK, ROBERT R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-442A  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-988-819-5

Query Match 97.3%; Score 1193; DB 3; Length 225;  
Best Local Similarity 96.9%; Pred. No. 1.8e-125;

Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDILTRYLEHQLRSLAGT 60  
Db 1 MDLRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDILTRYLEHQLRSLAGT 60

Qy 61 YLNYLGPPFPNEPDPNPRIGASTLPATVDLEVWRSNDKRLRNTQNYEAYSHLLCYLRGL 120  
Db 61 YLNYLGPPFPNEPDPNPRIGASTLPATVDLEVWRSNDKRLRNTQNYEAYSHLLCYLRGL 120

Qy 121 NROATAELRSLAHFCTSLQGLSIGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180  
Db 121 NROATAELRSLAHFCTSLQGLSIGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225  
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225

RESULT 7  
US-09-016-534-5  
; Sequence 5, Application US/09016534  
; Patent No. 6143874  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; APPLICANT: ELLIOTT, GARY S.  
; APPLICANT: SARMIENTO, ULLA  
; APPLICANT: SENALDI, GIORGIO  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/09/016,534  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442B  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-016-534-5

Query Match 97.3%; Score 1193; DB 4; Length 225;  
Best Local Similarity 96.9%; Pred. No. 1.8e-125;  
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDILTRYLEHQLRSLAGT 60  
Db 1 MDLRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDILTRYLEHQLRSLAGT 60

Qy 61 YLNYLGPPFPNEPDPNPRIGASTLPATVDLEVWRSNDKRLRNTQNYEAYSHLLCYLRGL 120  
Db 61 YLNYLGPPFPNEPDPNPRIGASTLPATVDLEVWRSNDKRLRNTQNYEAYSHLLCYLRGL 120

Qy 121 NROATAELRSLAHFCTSLQGLSIGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180  
Db 121 NROATAELRSLAHFCTSLQGLSIGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225  
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225

RESULT 8  
US-09-106-182-3  
; Sequence 3, Application US/09106182  
; Patent No. 6046035  
; GENERAL INFORMATION:  
; APPLICANT: Shi, Yangu  
; APPLICANT: Ruben, Steve  
; TITLE OF INVENTION: Cardiostrophin-Like Cytokine  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc  
; STREET: 9410 Key West Ave  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: US  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/09/106,182  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/051,053  
; FILING DATE: 30-JUN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF385  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 203 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-106-182-3

Query Match 10.2%; Score 124.5; DB 3; Length 203;  
Best Local Similarity 28.1%; Pred. No. 5.5e-06;  
Matches 50; Conservative 24; Mismatches 85; Indels 19; Gaps 5;

Qy 40 IQKTYDLTRYLEHQLRSLAGTYLVNLPFPNPFNPPRL---GAETLPRAVTDLVWRS 96  
Db 27 IRQTHNLRLLTKYADQLLEEVVQQGEPFGLPGFSPRLPLAGLSGPAPSHAGLPV--- 83

Qy 97 LNDKRLRNTQNYEAYSHLLCYLRGLNRQAA-----TAEIARRSLAHFCTSLQGLSIGVMA 151  
Db 84 ---SERLRQDAALASALPALLDAVRRRQALNPRAPRLRSLRLEDAARQVRALGAAVETVL 140

Qy 152 AALGY----PLPQPLPGTPTWTPGPAHSDFLQKMDDFWLLKELQTLWRSKDFNRL 205  
Db 141 AALGAAAGPVPEPV-ATSLFTSNSAAGVFSKVLGLHVCGLYGEWVSRTGDLGQL 197

RESULT 9  
US-08-233-609-3  
; Sequence 3, Application US/08233609







ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: 894PID2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-8674  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 203 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-443-129-3

[illegible]

RESULT 14  
US-08-443-952-3  
Sequence 3, Application US/08443952  
Patent No. 5679545  
GENERAL INFORMATION:  
APPLICANT: Baker, Joffre  
APPLICANT: Chien, Kenneth  
APPLICANT: King, Kathleen  
APPLICANT: Pennica, Diane  
APPLICANT: Wood, William  
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/443,952  
FILING DATE: 17-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/233609  
FILING DATE: 25-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/286304  
FILING DATE: 05-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: 894P1D4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-8674  
TELEFAX: 415/952-9881

```

; TELE: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-443-952-3

Query Match          9.7%; Score 118.5; DB 1; Length 203;
Best local similarity 28.8%; Pred. No. 2.6e-05;
Matches 51; Conservative 21; Mismatches 88; Indels 17; Gaps 5;

Qy 40 IOKTYDLTRYLEHQRLSLACTYLNLYLPPNPDNFNPRLL---CAETLPRATVDLEVWRS 96
Db 27 IRQTHNARLLTKVAEQLLREYVQQQGEFGLGCFSPRLPLAGLSGAPSPSHAGLPV--- 83
Qy 97 LNDKRLRTQNYEAYSHLLCYLRLGRLNQQAA-----TAELRSLAHFTSLQGLGSIAGVM 151
Db 84 ---SERLQRQAAALSVLPALLDAVRRRQAEINPRAPRLRLSLRSLSDAARQVRLGAAVETVL 140
Qy 152 AALGYPL--PQPLPGTEPT--WTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKQPNRL 205
Db 141 AALGAAARGPGEVTVATLFTANSTAGIISAKVLGHVCGLYGEWVSRTEGDLGQL 197

RESULT 15
US-08-443-130-3
; Sequence 3, Application US/08443130
; Patent No. 5723585
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,130
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/233609
; FILING DATE: 25-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/286304
; FILING DATE: 05-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: 894PID3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/325-8674
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

US-08-443-130-3

Query Match 9.7%; Score 118.5; DB 1; Length 203;  
Best Local Similarity 28.8%; Pred. No. 2.6e-05;  
Matches 51; Conservative 21; Mismatches 88; Indels 17; Gaps 5;  
Qy 40 IQTYDLTRYLEHQLRSLAGTYLNYLGGPPFNEPDNPPRL---GAETLPRATVDLEVWRS 96  
Db 27 IROTHNLARLLTKYAEQLLEEVVQQGEPFGLPGFSPPLPLAGLSGAPSHAGLPV--- 83  
Qy 97 LNDKLRLTQNVAYSHLLCYLRGLNRQAA-----TAE LRSLAHFCTSLQGLLSIAGVM 151  
Db 84 ---SERLRQDAALSVLPALLDAVRRQAEINPRAPRLRLRSLEDAARQVRALGAAVETVL 140  
Qy 152 AALGYPL--PQPLFGTEPT-WTPGPAHSDFLQKMDDFWLLKELQTLWRSAKDFNRL 205  
Db 141 AALGAAARGPGPEPVTATLFTANSTAGIFSAKVLGFHVCGLYGEWVSRTEGDLGQL 197

Search completed: January 27, 2003, 15:43:07  
Job time : 14 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 12532.7 Seconds  
(without alignments)  
11812.771 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 5087

Sequence: 1 aacctggagtggtggcgtggc.....cctttgtaagcttcctca 5087

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pi.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	5087	100.0	5087	6	AR002596	Sequence
2	5087	100.0	5087	6	AX392088	Sequence
3	5087	100.0	5087	9	AF176912	Homo sapi
C	4910.4	96.5	168567	9	AP003419	Homo sapi
C	4908.8	96.5	169144	2	AC005849	Homo sapi
5	4605	90.5	135116	2	AF002437	Homo sapi
7	1709.4	33.6	269355	2	AC109138	Homo sapi
8	1463.4	28.8	1736	9	BC012939	Mus muscu
9	1461.4	28.7	1689	9	AF172854	Homo sapi
10	525.4	10.3	881	6	AX205024	Sequence
11	525.4	10.3	881	6	AX205042	Sequence
12	523.4	10.3	797	6	AR002595	Sequence
13	523.4	10.3	797	6	AX392086	Sequence
14	523.4	10.3	797	9	AF176911	Sequence
15	495.4	9.7	680	9	AY049779	Homo sapi
16	492.8	9.7	1692	6	AX205060	Sequence
17	429.6	8.4	819	6	AR002597	Sequence
18	429.6	8.4	819	6	AX392089	Sequence
19	429.6	8.4	819	10	AF176913	Sequence
20	165.2	3.2	283	6	AX202145	Mus muscu
C	148.8	2.9	63347	2	AC110526	Sequence
22	99	1.9	7218	6	I66494	Sequence 14
23	82.8	1.6	179145	9	AC009785	Homo sapi
24	78	1.5	257393	2	AC102125	Homo sapi
25	73.8	1.5	203982	9	AC067854	Homo sapi
C	73.2	1.4	195690	10	AC098726	Mus muscu
26	72	1.4	98179	2	AC117037	Rattus no
C	70.6	1.4	95097	9	AL133232	Human DNA
28	69.4	1.4	196876	2	AC097558	Rattus no
29	68.8	1.4	35848	2	AC011553	Homo sapi
C	68.8	1.4	125020	9	AF429315	Homo sapi
31	68.2	1.3	161984	2	AC115241	Rattus no
C	67.6	1.3	10732	6	E32986	Gene encodi
33	67.6	1.3	163246	2	AC068401	Homo sapi
34	67.6	1.3	174428	9	AC005821	Homo sapi
35	67.6	1.3	125020	9	AF429315	Homo sapi
C	67.2	1.3	125020	9	AF429315	Homo sapi
37	67	1.3	104518	2	AC130185	Ornithorh
C	66.8	1.3	230000	9	AF243527	Homo sapi
38	65.8	1.3	230687	2	AC124433	Mus muscu
C	65.6	1.3	200947	2	AC016961	Homo sapi
40	65	1.3	171624	2	AC095621	Rattus no
C	65	1.3	196283	10	AL672215	Mouse DNA
42	65	1.3	321438	2	AC104989	Homo sapi
C	64.8	1.3	139729	2	AC104100	Mus muscu
44	64.6	1.3	109975	9	AC112497	Homo sapi
C	64.6	1.3	109975	9	AC112497	Homo sapi

ALIGNMENTS

RESULT 1  
AR002596  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES

AR002596  
Sequence 3 from patent US 5741772.  
AR002596  
AR002596.1 GI:3964150  
Unknown.  
Unknown.  
Unclassified.  
1 (bases 1 to 5087)  
Chang,M.-s.  
Neurotrophic factor NNT-1  
Patent: US 5741772-A 3 21-APR-1998;  
Location/Qualifiers

5087 bp  
DNA  
linear  
PAT 04-DEC-1998

[illegible]

Db	2041		CAGTAGGCGAGGGGTTTAGGTTGGATGGAGTTATGCTTAAAGATTGGGGGTGG	2100
Qy	2101		TCCAGAGGTGTTACAGAGGCCAGGAGAGGAAGAGGGTTGGAGGAGCCGAGGCACC	2160
Db	2101		TCCAGAGGTGTTACAGAGGCCAGGAGAGGAAGAGGGTTGGAGGAGCCGAGGCACC	2160
Qy	2161		ATGGGGAAACGGGCCCTCTTCCGGTTCCTCTTCCACATCCAGACCCCTACTCTGGAG	2220
Db	2161		ATGGGGAAACGGGCCCTCTTCCGGTTCCTCTTCCACATCCAGACCCCTACTCTGGAG	2220
Qy	2221		CCAGGAAAGAAAGGAGAGGTGGGGGGAGCTGGCTCAGCCCCCAGGATACACCG	2280
Db	2221		CCAGGAAAGAAAGGAGAGGTGGGGGGAGCTGGCTCAGCCCCCAGGATACACCG	2280
Qy	2281		AGGAAATTAGTTTGTCTCTGTGTTGCAGGTGTGAACCTCCGCCCTGGGCCCTTGCGCTA	2340
Db	2281		AGGAAATTAGTTTGTCTCTGTGTTGCAGGTGTGAACCTCCGCCCTGGGCCCTTGCGCTA	2340
Qy	2341		TCCAGGCTCTCCCTTGCTTCTCCGTTCTTTCCAGTTATACATCTCCCTCATCCCTT	2400
Db	2341		TCCAGGCTCTCCCTTGCTTCTCCGTTCTTTCCAGTTATACATCTCCCTCATCCCTT	2400
Qy	2401		TCCCTGGCCCCAGCGCTCCCGGAGGTTGAAAGGGCTCTGGCCTCTTCCTATACC	2460
Db	2401		TCCCTGGCCCCAGCGCTCCCGGAGGTTGAAAGGGCTCTGGCCTCTTCCTATACC	2460
Qy	2461		ATGCTGTCTTCCATAGCTTCCCTGTCTACTCATAGAGCTGCTCCATTTCTTCCCTT	2520
Db	2461		ATGCTGTCTTCCATAGCTTCCCTGTCTACTCATAGAGCTGCTCCATTTCTTCCCTT	2520
Qy	2521		CTGCAACCTGTCTTATCAGCTGAACCCCTTTCTTCGGAGTGTAGTGAGTACCCGTCTC	2580
Db	2521		CTGCAACCTGTCTTATCAGCTGAACCCCTTTCTTCGGAGTGTAGTGAGTACCCGTCTC	2580
Qy	2581		TCCCCAGCCCTCAGCTGGTGGCCTGGGTGTGTGAGCGGCAATGGGGCTCTGGTTCCA	2640
Db	2581		TCCCCAGCCCTCAGCTGGTGGCCTGGGTGTGTGAGCGGCAATGGGGCTCTGGTTCCA	2640
Qy	2641		ATGGGCCACTCTCATCTCTCTTGTTCCTTGTGCAAGAAACCTTTGCTTCACTCCACATG	2700
Db	2641		ATGGGCCACTCTCATCTCTCTCTTGTTCCTTGTGCAAGAAACCTTTGCTTCACTCCACATG	2700
Qy	2701		CCCTCTCTAGTTCCGACCCCTTTTCTCTGGCTTCCCTGCCAAATTTCTCCAAGGA	2760
Db	2701		CCCTCTCTAGTTCCGACCCCTTTTCTCTGGCTTCCCTGCCAAATTTCTCCAAGGA	2760
Qy	2761		GTGGTCTACACCTCTGCTCCACTTCTCTCCACCCACTCACTCTTAAACCCCTGAA	2820
Db	2761		GTGGTCTACACCTCTGCTCCACTTCTCTCTCCACCCACTCACTCTTAAACCCCTGAA	2820
Qy	2821		TCTGGCTTCCAGGCCCCAGCAATGGTTCTCTCAAAGTCTCAGGCACCTCTTGCCAAAG	2880
Db	2821		TCTGGCTTCCAGGCCCCAGCAATGGTTCTCTCAAAGTCTCAGGCACCTCTTGCCAAAG	2880
Qy	2881		CCCGACAGTGTTTGAAAGCTCAITCTCTCTGGTCTGTGTTTGAGCCACACTGCTGAG	2940
Db	2881		CCCGACAGTGTTTGAAAGCTCAITCTCTCTGGTCTGTGTTTGAGCCACACTGCTGAG	2940
Qy	2941		CGCTCTGCTTCTCGAACTCTCTTCTGGTCTGTGACTCTCTGGGCCACCTTCTA	3000
Db	2941		CGCTCTGCTTCTCGAACTCTCTTCTGGTCTGTGACTCTCTGGGCCACCTTCTA	3000
Qy	3001		CCTCTCAGCTCTCCAGGCTCCTTCTCTCTGCTGTGCTGCCCCACAGCGGGCACTCTC	3060
Db	3001		CCTCTCAGCTCTCCAGGCTCCTTCTCTCTGCTGTGCTGCCCCACAGCGGGCACTCTC	3060
Qy	3061		CCAGGTTTGGCCACCCAGCCCATCAGACGTCCTTCTGTGAGCGTCTTGTGCGTCTCCTC	3120
Db	3061		CCAGGTTTGGCCACCCAGCCCATCAGACGTCCTTCTGTGAGCGTCTTGTGCGTCTCCTC	3120
Qy	3121		CTCTCTCTTTTCTACGCTCTCCATTTGGAGCTCACCAAGCCACTGCTTCAACTGTC	3180

Db	3121	CTCTCTCTTTTCTTAGCCCTCTCCATTGGAGAGCTCACCCGCGCACTGCTTCAACTGTCTC	3180
QY	3181	ACCTGCATACAAATGATATCTTTATTGAAAAAATCAGGAGGCGCATGAACAAGAAGACC	3240
Db	3181	ACCTGCATACAAATGATATCTTTATTGAAAAAATCAGGAGGCGCATGAACAAGAAGACC	3240
QY	3241	TAGCATGGAGACAGGGCCAGTGTTCAGGGGACACAAAAATAGAAAACTTTGGGAGCAGGTA	3300
Db	3241	TAGCATGGAGACAGGGCCAGTGTTCAGGGGACACAAAAATAGAAAACTTTGGGAGCAGGTA	3300
QY	3301	TCCTCTTGGTGTGAGCCAGCGCTCTGCGCTCTCTCTCTCCCATCACCTCTCTCTTTTC	3360
Db	3301	TCCTCTTGGTGTGAGCCAGCGCTCTGCGCTCTCTCTCTCCCATCACCTCTCTCTTTTC	3360
QY	3361	ACAGCTGAACCTACTCTGGGCCCCCTTTCAACAGCGCAGACTTCAACCCCTCCCGCCTGGG	3420
Db	3361	ACAGCTGAACCTACTCTGGGCCCCCTTTCAACAGCGCAGACTTCAACCCCTCCCGCCTGGG	3420
QY	3421	GGCAGAGACTCTGCCCAGGGCCACTGTGACTTGGAGGTGTGGGAGGCTCAATGACAA	3480
Db	3421	GGCAGAGACTCTGCCCAGGGCCACTGTGACTTGGAGGTGTGGGAGGCTCAATGACAA	3480
QY	3481	ACTGCGGCTGAACCCAGAACTTACGAGGCCCTACAGCCACCTTCTGTGTACTTGGCGTGGCCT	3540
Db	3481	ACTGCGGCTGAACCCAGAACTTACGAGGCCCTACAGCCACCTTCTGTGTACTTGGCGTGGCCT	3540
QY	3541	CAACCGTCAAGCTGCCACCTGCTGAGCTGCGCGCAGCTCTGGGCTACCCACTGCCCCA	3600
Db	3541	CAACCGTCAAGCTGCCACCTGCTGAGCTGCGCGCAGCTCTGGGCTACCCACTGCCCCA	3600
QY	3601	CCAGGCGCTGCTGGGCGAGCAITTCGGGCGCTCATGGCAGCTCTGGGCTACCCACTGCCCCA	3660
Db	3601	CCAGGCGCTGCTGGGCGAGCAITTCGGGCGCTCATGGCAGCTCTGGGCTACCCACTGCCCCA	3660
QY	3661	GGCGTCTGGGACTGAACCCACTTTGGAATCTCTGGCCCTGCCCAcAGTGAATCTCTCCA	3720
Db	3661	GGCGTCTGGGACTGAACCCACTTTGGAATCTCTGGCCCTGCCCAcAGTGAATCTCTCCA	3720
QY	3721	GAAGATGGAACGACTTCTGGCTGCTGAAGAGCTGCAGACCTGGCTGTGGCGTTCGGCCAA	3780
Db	3721	GAAGATGGAACGACTTCTGGCTGCTGAAGAGCTGCAGACCTGGCTGTGGCGTTCGGCCAA	3780
QY	3781	GSACTTCAACCGCTCAGAAGAGATGAGCGCTCAGCAGTGCAGTCAACCTGCACCT	3840
Db	3781	GSACTTCAACCGGCTCAAGAAAGATGAGCGCTCAGCAGTGCAGTCAACCTGCACCT	3840
QY	3841	GGGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCTTCGCTCCCTTCACACCTGCT	3900
Db	3841	GGGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCTTCGCTCCCTTCACACCTGCT	3900
QY	3901	CCCACTTTGTGAGACCGCCTGATGCCAACACTGTTGAGCCAGGAGACAGAAGCTG	3960
Db	3901	CCCACTTTGTGAGACCGCCTGATGCCAACACTGTTGAGCCAGGAGACAGAAGCTG	3960
QY	3961	TGAGCTCTGGCCCTTCTCGACCGGCTGGCGGTGTATGCGATCAGCCCTGTCTCTC	4020
Db	3961	TGAGCTCTGGCCCTTCTCGACCGGCTGGCGGTGTATGCGATCAGCCCTGTCTCTC	4020
QY	4021	CCCACTCCCAAGGCTTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTGTCTG	4080
Db	4021	CCCACTCCCAAGGCTTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTGTCTG	4080
QY	4081	TTTCTACAGGAAGTCATGCTCGAGGAGTGTGAATGTTTCAAGTTGTGTGAGAGCGCT	4140
Db	4081	TTTCTACAGGAAGTCATGCTCGAGGAGTGTGAAGTTGTGTGAGAGCGCT	4140
QY	4141	CATGCCCTCTGCTTCTTGCCCTACCACTTGGCCAGTGCCCAcCAGCCCTCAGGTGCA	4200
Db	4141	CATGCCCTCTGCTTCTTGCCCTACCACTTGGCCAGTGCCCAcCAGCCCTCAGGTGCA	4200
QY	4201	CATCTGAGGCGCAGGGTTGAGGGGCCACCAcCAcATGCTTTCTGGGCTGAAGCCT	4260
Db	4201	CATCTGAGGCGCAGGGTTGAGGGGCCACCAcCAcATGCTTTCTGGGCTGAAGCCT	4260

Qy	4261	TTGGCTGCCCACTCTCTTGAGAGGGTGTGCTCCCTTATCCCAATCACTCTATACA	4320
Db	4261	TTGGCTGCCCACTCTCTTGAGAGGGTGTGCTCCCTTATCCCAATCACTCTATACA	4320
Qy	4321	TCCATTTCAGGAAACAAACATGGTGCATTTCTACACAAAAGAGATGAGATTAAACAGTG	4380
Db	4321	TCCATTTCAGGAAACAAACATGGTGCATTTCTACACAAAAGAGATGAGATTAAACAGTG	4380
Qy	4381	CAGGGTTCGGGTCTGCATTTGGAGGTCCCTTATAAAACGAGAGAAAATACTGAAAGCAC	4440
Db	4381	CAGGGTTCGGGTCTGCATTTGGAGGTCCCTTATAAAACGAGAGAAAATACTGAAAGCAC	4440
Qy	4441	AGGGCAGGGAACAGACACAGACCCAGGAGTCTCCAAAGCACAGAGTGGCAACAAA	4500
Db	4441	AGGGCAGGGAACAGACACAGACCCAGGAGTCTCCAAAGCACAGAGTGGCAACAAA	4500
Qy	4501	ACCCGAGCTGAGCATCAGACACCTTGCTCGAATTGCTTCCAGTATTACGGTGCCTCTTC	4560
Db	4501	ACCCGAGCTGAGCATCAGACACCTTGCTCGAATTGCTTCCAGTATTACGGTGCCTCTTC	4560
Qy	4561	TCCTGCCCTCTTCCAGGGTATCTGTGGTTGCCAGCTGGGAGGCAACCATAGCCAC	4620
Db	4561	TCCTGCCCTCTTCCAGGGTATCTGTGGTTGCCAGCTGGGAGGCAACCATAGCCAC	4620
Qy	4621	ACCAGAGATTCTGAAAGTTTACATGCACTAGCAATTTTGGGTGTAGGTGGCAGCT	4680
Db	4621	ACCAGAGATTCTGAAAGTTTACATGCACTAGCAATTTTGGGTGTAGGTGGCAGCT	4680
Qy	4681	CCCCAAGGCTCTGCCCCAGCCCAACCCACTCATGACTCTAAGTGTGTTGTAATAT	4740
Db	4681	CCCCAAGGCTCTGCCCCAGCCCAACCCACTCATGACTCTAAGTGTGTTGTAATAT	4740
Qy	4741	TTATTTATTGGAGATGTTATTATTATAGATGATATTTATTGCGAATTTCTATTCTTGTA	4800
Db	4741	TTATTTATTGGAGATGTTATTATTATAGATGATATTTATTGCGAATTTCTATTCTTGTA	4800
Qy	4801	TTAACAATAAATGCTTGCCCAAGAACTAGTCTCTTTGGCCAGCTCACCCCTCTCTGG	4860
Db	4801	TTAACAATAAATGCTTGCCCAAGAACTAGTCTCTTTGGCCAGCTCACCCCTCTCTGG	4860
Qy	4861	TGCTCATCAGACTCTTGCCACCCCTGGCTCCCACTCCCTGCTTCCCTGTGGTGGAGTGC	4920
Db	4861	TGCTCATCAGACTCTTGCCACCCCTGGCTCCCACTCCCTGCTTCCCTGTGGTGGAGTGC	4920
Qy	4921	ACAGAGCTCTGGAGAGGCGCTTCTCTCCCGCACTGGGGCGATGGGGCGACCTCAGA	4980
Db	4921	ACAGAGCTCTGGAGAGGCGCTTCTCTCCCGCACTGGGGCGATGGGGCGACCTCAGA	4980
Qy	4981	CTTACCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCTCCACACAGCTTC	5040
Db	4981	CTTACCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCTCCACACAGCTTC	5040
Qy	5041	TGTCACCCAGGTTTCCCTCACCCCACTTTGCTAAGTCTTCCTCA	5087
Db	5041	TGTCACCCAGGTTTCCCTCACCCCACTTTGCTAAGTCTTCCTCA	5087
RESULT 2			
LOCUS	AX392088	5087 bp	DNA linear PAT 23-MAR-2002
DEFINITION	Sequence 3 from Patent WO0215977.		
ACCESSION	AX392088		
VERSION	AX392088.1	GI:19700576	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	Senaldi, G.		
AUTHORS	Methods and compositions for treating ige-related disease using nn		
TITLE	t-i inhibitors		

JOURNAL Patent: WO 0215977-A 3 28-FEB-2002;

Amgen Inc. (US)

FEATURES Location/Qualifiers

source 1..5087

/organism="Homo sapiens"

/db\_xref="taxon:9606"

misc\_feature 137..138

/note="INTERVENING UNSEQUENCED REGION OF >1KB"

BASE COUNT 992 a 1746 c 1191 g 1158 t

ORIGIN

Query Match 100.0%; Score 5087; DB 6; Length 5087;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACCTGGAGTGGGCTGGGGATGGGATTATTAAAGCTTCGCCGGAGCCGGCTCGCC 60

Db 1 AACCTGGAGTGGGCTGGGGATGGGATTATTAAAGCTTCGCCGGAGCCGGCTCGCC 60

Qy 61 CTCCTCACTCCGCCAGCTCCGGGAGAGGAGCGCACCCGGCGGCCAGCCGCCA 120

Db 61 CTCCTCACTCCGCCAGCTCCGGGAGAGGAGCGCGCACCCGGCGGCCAGCCGCCA 120

Qy 121 TGGACCTCCGAGAGGTTGAAAACCCAACTAGCCCTGCTTTCATATACATGACAGCAG 180

Db 121 TGGACCTCCGAGAGGTTGAAAACCCAACTAGCCCTGCTTTCATATACATGACAGCAG 180

Qy 181 CGCCCCATCTGATACCTAAACCGCACCAAGTCAAGCCCTCCAACTCACCTCTGCTGCC 240

Db 181 CGCCCCATCTGATACCTAAACCGCACCAAGTCAAGCCCTCCAACTCACCTCTGCTGCC 240

Qy 241 CAGACCTCACCACATCTTTGGAGTCAAACTCAACCGCACTAAATCAACCAATCCCA 300

Db 241 CAGACCTCACCACATCTTTGGAGTCAAACTCAACCGCACTAAATCAACCAATCCCA 300

Qy 301 AGCTAAACTAATCTGAAACTTTTAAAGTAAACCCAGTCTTAAACCTAACCTAGCCCAAT 360

Db 301 AGCTAAACTAATCTGAAACTTTTAAAGTAAACCCAGTCTTAAACCTAACCTAGCCCAAT 360

Qy 361 GCCAATTATCTACCTAGCCAAACCCCTAACTGCTTTGCCAGTCCAAAGTGTCCACTG 420

Db 361 GCCAATTATCTACCTAGCCAAACCCCTAACTGCTTTGCCAGTCCAAAGTGTCCACTG 420

Qy 421 AATCTCACTCTGGTCTCACTGAAAATCCAGAAAAGCATATTTCCCACTGCCACAT 480

Db 421 AATCTCACTCTGGTCTCACTGAAAATCCAGAAAAGCATATTTCCCACTGCCACAT 480

Qy 481 CCCTCTTACAGCACCCCAACCTGCGCTCTGCACTCTGCTGATCTCTGGATGTCCAACT 540

Db 481 CCCTCTTACAGCACCCCAACCTGCGCTCTGCACTCTGCTGATCTCTGGATGTCCAACT 540

Qy 541 CTGCACTGCGCATCAGCAACAAAGCCGACTCGTCAAAATGACCTCTCTCTCTCTCTCC 600

Db 541 CTGCACTGCGCATCAGCAACAAAGCCGACTCGTCAAAATGACCTCTCTCTCTCTCTCC 600

Qy 601 CCACCTCTGAGCTGATGAAAAGGCTCATTTGAAGTCCAACTTTCCCACTTAAACACC 660

Db 601 CCACCTCTGAGCTGATGAAAAGGCTCATTTGAAGTCCAACTTTCCCACTTAAACACC 660

Qy 661 AAGACGGGGTGAACCTCCACACTGCGCCCTGAGAGTGAGGAGTAAATCTCTCT 720

Db 661 AAGACGGGGTGAACCTCCACACTGCGCCCTGAGAGTGAGGAGTAAATCTCTCTCT 720

Qy 721 CAATCTAACCCCAACCTTACACTTCCCACTCAGGAATCATCTTGAATATATACCCAAA 780

Db 721 CAATCTAACCCCAACCTTACACTTCCCACTCAGGAATCATCTTGAATATATACCCAAA 780

Qy 781 ACTAAGCCCCATAGGAGCGCCGAGCTAGTGTCTAAACCTTATACCTTCTCTCTATGG 840

Db 781 ACTAAGCCCCATAGGAGCGCCGAGCTAGTGTCTAAACCTTATACCTTCTCTCTATGG 840

Qy 841 GTGAGTCTGTTCTGGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900

Db 841 GTGAGTCTGTTCTGGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900

Db 841 GTGAGTCTGTTTGGGGCGCGCTCTCTCTCTGTTCTCCCTTAGAGCTGACTGTGCTC 900  
Qy 901 AGCTGCGAGCTCTGACATGCTGTCTCCCAACCTCTGACTCCCTCAAGCTCAGTGG 960  
Db 901 AGCTGCGAGCTCTGACATGCTGTCTCCCAACCTCTGACTCCCTCAAGCTCAGTGG 960  
Qy 961 GACTGGAGACTGCGAGAGCTAGGGTAACTCTGGAACAAGGAGGCTGAGT 1020  
Db 961 GACTGGAGACTGCGAGAGCTAGGGTAACTCTGGAACAAGGAGGCTGAGT 1020  
Qy 1021 CCCTAGGCTGGCCCGCTCCCTCATGCTGACATATACATGTTGGACACACAGTG 1080  
Db 1021 CCCTAGGCTGGCCCGCTCCCTCATGCTGACATATACATGTTGGACACACAGTG 1080  
Qy 1081 GCACATGCAAGACTCTCTGAGCTGACACAGATCCATTCTCAAGTATCTACTGAT 1140  
Db 1081 GCACATGCAAGACTCTCTGAGCTGACACAGATCCATTCTCAAGTATCTACTGAT 1140  
Qy 1141 AGACATGCTGCGCAAGTCTCATCTCTCAATCAATACATGCTCTCTCTCTCTCCC 1200  
Db 1141 AGACATGCTGCGCAAGTCTCATCTCTCAATCAATACATGCTCTCTCTCTCTCCC 1200  
Qy 1201 GTCTTGGCAGAGTGTTCCTCCCTCCATCCCTCTGCTCCCTCCCTCCCTCCCTCC 1260  
Db 1201 GTCTTGGCAGAGTGTTCCTCCCTCCATCCCTCTGCTCCCTCCCTCCCTCCCTCC 1260  
Qy 1261 TCACCCCCACCCAGGCTGGGACAGACACTGAGGGCTGCCAGCTGCTCTCCC 1320  
Db 1261 TCACCCCCACCCAGGCTGGGACAGACACTGAGGGCTGCCAGCTGCTCTCCC 1320  
Qy 1321 GTGTGGGCGGGCGGCTCATGCTTCTGCTCCATCTGCTCCCAAGGGACTCGTGGG 1380  
Db 1321 GTGTGGGCGGGCGGCTCATGCTTCTGCTCCATCTGCTCCCAAGGGACTCGTGGG 1380  
Qy 1381 GATGTTAGCTGCTGTGACAGTGTCTGACACTCTGCTCCCTGCTGCTGCTGCTGCT 1440  
Db 1381 GATGTTAGCTGCTGTGACAGTGTCTGACACTCTGCTCCCTGCTGCTGCTGCTGCT 1440  
Qy 1441 CACAGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCCGCTACCT 1500  
Db 1441 CACAGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCCGCTACCT 1500  
Qy 1501 GGAGCAACAACTCCGCAAGTGGTGGGACCTATGTGATATCAGCGTAGGAATCTGGG 1560  
Db 1501 GGAGCAACAACTCCGCAAGTGGTGGGACCTATGTGATATCAGCGTAGGAATCTGGG 1560  
Qy 1561 AGTTGGGAGGAGTGAGAGTTGGGAAAGACAGTCTTAACCTGGAGGTTCTGGTAAA 1620  
Db 1561 AGTTGGGAGGAGTGAGAGTTGGGAAAGACAGTCTTAACCTGGAGGTTCTGGTAAA 1620  
Qy 1621 TGATGGGTTGAGGAGGCTCTTTGGCTCCACAGTCCCTGCTGCTGCTATCTCTG 1680  
Db 1621 TGATGGGTTGAGGAGGCTCTTTGGCTCCACAGTCCCTGCTGCTGCTATCTCTG 1680  
Qy 1681 CCCTTCCCTCTTAGGTGGCCCCCCTATCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740  
Db 1681 CCCTTCCCTCTTAGGTGGCCCCCCTATCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740  
Qy 1741 GGCAGGCTCGCACCGGCTGGGCCATATGGCCCACTGGCTGCGACGCCAGCGCGCGCC 1800  
Db 1741 GGCAGGCTCGCACCGGCTGGGCCATATGGCCCACTGGCTGCGACGCCAGCGCGCGCC 1800  
Qy 1801 TCCCTCTGGGGCGGGAGTCTCTCTGTTTACCGTGTGTGTTGTTGTTGTTGTTGTTG 1860  
Db 1801 TCCCTCTGGGGCGGGAGTCTCTCTGTTTACCGTGTGTGTTGTTGTTGTTGTTGTTG 1860  
Qy 1861 GGCAGGTTGGGTTGGGACAGAGGGGCCCACTCCCAATGCTGCTGCTGCTGCTGCTGCT 1920  
Db 1861 GGCAGGTTGGGTTGGGACAGAGGGGCCCACTCCCAATGCTGCTGCTGCTGCTGCTGCT 1920  
Qy 1921 TGCCCCCAGACTGGGGCTGTGCTGTGACCCAGGGGCTCCCTTTCGCTGCTGCTGCT 1980  
Db 1921 TGCCCCCAGACTGGGGCTGTGCTGTGACCCAGGGGCTCCCTTTCGCTGCTGCTGCT 1980

Qy 1981 CCCATCTAGTGGGCCCTCCTAGGGGGTCACTGGGGGAAGGGGACTGTAGGGAACCCAGG 2040  
Db 1981 CCCATCTAGTGGGCCCTCCTAGGGGGTCACTGGGGGAAGGGGACTGTAGGGAACCCAGG 2040  
Qy 2041 CAGTAGTGGCAGGGGGTTAGGGTGTGATGAGAGTTATGCTGTAAAGATTGGGGGTGG 2100  
Db 2041 CAGTAGTGGCAGGGGGTTAGGGTGTGATGAGAGTTATGCTGTAAAGATTGGGGGTGG 2100  
Qy 2101 TCCAGAGGTGTTCAAGAGAGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
Db 2101 TCCAGAGGTGTTCAAGAGAGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
Qy 2161 ATGGGGAACCGGCCCTCTTCCCGTGTCTCTTCCACATCCAGAGCCCTACTCTGGAG 2220  
Db 2161 ATGGGGAACCGGCCCTCTTCCCGTGTCTCTTCCACATCCAGAGCCCTACTCTGGAG 2220  
Qy 2221 CCAGGGAAGAAAG 2280  
Db 2221 CCAGGGAAGAAAG 2280  
Qy 2281 AGGAAATTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2340  
Db 2281 AGGAAATTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2340  
Qy 2341 TCCAGAGCTCTCCCGTGTCTCTTCCCTTCTTCCAGATTATACATCTCCCTCATCCCTT 2400  
Db 2341 TCCAGAGCTCTCCCGTGTCTCTTCCCTTCTTCCAGATTATACATCTCCCTCATCCCTT 2400  
Qy 2401 TCCCTGGGCGCCAGCGCTCCCGAGAGGTTGAAAGAGGCTCTGCTCTCTCTCTCTCT 2460  
Db 2401 TCCCTGGGCGCCAGCGCTCCCGAGAGGTTGAAAGAGGCTCTGCTCTCTCTCTCTCT 2460  
Qy 2461 ATGCTGTTTCCATAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520  
Db 2461 ATGCTGTTTCCATAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520  
Qy 2521 CTGCAACCT 2580  
Db 2521 CTGCAACCT 2580  
Qy 2581 TCCCGAGCCCTCAGCTGGGGCTGGGTGTGTGAGCGGCAATGGGGCTCTGCTTCCA 2640  
Db 2581 TCCCGAGCCCTCAGCTGGGGCTGGGTGTGTGAGCGGCAATGGGGCTCTGCTTCCA 2640  
Qy 2641 ATGGGCACTCTCATCT 2700  
Db 2641 ATGGGCACTCTCATCT 2700  
Qy 2701 CCCTCTCTAGTTCCCGAGCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760  
Db 2701 CCCTCTCTAGTTCCCGAGCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760  
Qy 2761 GTGCTCTACACCTCTGCT 2820  
Db 2761 GTGCTCTACACCTCTGCT 2820  
Qy 2821 TCTGGCTTCCAGGCCCGCAGCAATGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880  
Db 2821 TCTGGCTTCCAGGCCCGCAGCAATGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880  
Qy 2881 CCCGACAGTGTGTTTGAAGGCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940  
Db 2881 CCCGACAGTGTGTTTGAAGGCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940  
Qy 2941 CGCTGCT 3000  
Db 2941 CGCTGCT 3000  
Qy 3001 CCTCTCAGCT 3060  
Db 3001 CCTCTCAGCT 3060

Qy 3061 CCAAGTTTGGCCACCCAGCCCAATACAGCAGTCCTTCTTGAGCGTCTTGTCGCTCTCCTC 3120  
Db 3061 CCAAGTTTGGCCACCCAGCCCAATACAGCAGTCCTTCTTGAGCGTCTTGTCGCTCTCCTC 3120  
Qy 3121 CTCCTCCCTTTTCTAGGCTCTCCATTGGAGAGCTCACACCGCCACTGCTTCAACTGTC 3180  
Db 3121 CTCCTCCCTTTTCTAGGCTCTCCATTGGAGAGCTCACACCGCCACTGCTTCAACTGTC 3180  
Qy 3181 ACCTGCATCAAAATGATATCTTATTGGAAAACTCAGGGAGGCCATGAACAAAGAGGCC 3240  
Db 3181 ACCTGCATCAAAATGATATCTTATTGGAAAACTCAGGGAGGCCATGAACAAAGAGGCC 3240  
Qy 3241 TAGCATGGAGACAGGCGCCAGTGTACGGGACACAAAAATAGAAAACTTTGGGAGCAGGTA 3300  
Db 3241 TAGCATGGAGACAGGCGCCAGTGTACGGGACACAAAAATAGAAAACTTTGGGAGCAGGTA 3300  
Qy 3301 TCTCCTTGGTGTGAGCCAGCGGCTCTGCCCTCTCTTCCCTCATCACCCTCTCCTTTTC 3360  
Db 3301 TCTCCTTGGTGTGAGCCAGCGGCTCTGCCCTCTCTTCCCTCATCACCCTCTCCTTTTC 3360  
Qy 3361 ACAGCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGG 3420  
Db 3361 ACAGCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGG 3420  
Qy 3421 GGCAGAGACTCTGCCAGGGCCACTGTGTGACTTGGAGGTGTGGCGAAGCCCTCAATGACAA 3480  
Db 3421 GGCAGAGACTCTGCCAGGGCCACTGTGTGACTTGGAGGTGTGGCGAAGCCCTCAATGACAA 3480  
Qy 3481 ACTGGGCTGACCCAGAACTACGAGGCCCTACAGCCACTTCTGTGTACTTGGCGTGGCCT 3540  
Db 3481 ACTGGGCTGACCCAGAACTACGAGGCCCTACAGCCACTTCTGTGTACTTGGCGTGGCCT 3540  
Qy 3541 CAAACGCTCAGGCTGCCACTGTGAGCTGGCGCCAGCTTGGCCCACTTCTGACACCAAGCCCT 3600  
Db 3541 CAAACGCTCAGGCTGCCACTGTGAGCTGGCGCCAGCTTGGCCCACTTCTGACACCAAGCCCT 3600  
Qy 3601 CAGGGCTGTGGGAGCATTTGGCGGCTGATGGAGCTCTGGGCTACCCACTGCCCCA 3660  
Db 3601 CAGGGCTGTGGGAGCATTTGGCGGCTGATGGAGCTCTGGGCTACCCACTGCCCCA 3660  
Qy 3661 GCGGCTGCTGGAGCTGAACCCACTTGGACTCTGGCCCTGCCCCACAGTGACTTCTCCA 3720  
Db 3661 GCGGCTGCTGGAGCTGAACCCACTTGGACTCTGGCCCTGCCCCACAGTGACTTCTCCA 3720  
Qy 3721 GAAGATGGACGACTTCTGGCTGTGAAGGAGCTGCAGACTGGCTGTGGCGCTCGGCCAA 3780  
Db 3721 GAAGATGGACGACTTCTGGCTGTGAAGGAGCTGCAGACTGGCTGTGGCGCTCGGCCAA 3780  
Qy 3781 GGAATTCACCGGCTCAAGAAGATGAGAGCTCCAGAGCTGCGAGTCAACCTTGACCT 3840  
Db 3781 GGAATTCACCGGCTCAAGAAGATGAGAGCTCCAGAGCTGCGAGTCAACCTTGACCT 3840  
Qy 3841 GGGGGCTCATGGCTCTGACTCTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900  
Db 3841 GGGGGCTCATGGCTCTGACTCTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900  
Qy 3901 CCCACTTTGTGAGAGCCAGCCCTGTATGCAACACCTGTGTAGCCAGGAGACAGAGCTG 3960  
Db 3901 CCCACTTTGTGAGAGCCAGCCCTGTATGCAACACCTGTGTAGCCAGGAGACAGAGCTG 3960  
Qy 3961 TGAGCTCTGGCCCTTTCTTGACCGGCTGGGCGTGTGATGCGATCAGCCCTGTCTCCTC 4020  
Db 3961 TGAGCTCTGGCCCTTTCTTGACCGGCTGGGCGTGTGATGCGATCAGCCCTGTCTCCTC 4020  
Qy 4021 CCCACTCCCAAGTCTACCGAGCTGGGAGGAGTACAGTAGGCGCTGTCTCTCTCTCTCTCT 4080  
Db 4021 CCCACTCCCAAGTCTACCGAGCTGGGAGGAGTACAGTAGGCGCTGTCTCTCTCTCTCTCT 4080  
Qy 4081 TTTTACAGGAAGTCAATGCTCAGGGAGTGTGAAGTGGTTTTCAGTGGTGGTGCAGAGCGCT 4140  
Db 4081 TTTTACAGGAAGTCAATGCTCAGGGAGTGTGAAGTGGTTTTCAGTGGTGGTGCAGAGCGCT 4140  
Qy 4141 CATGGCTCTCTGCTTCTGCTTACCACTTGGCCAGTGGCCACCCAGCCCTCAGGTGGCA 4200

Db 4141 CATGGCTCTCTGCTTCTTGGCTTACCACTTGGCCAGTGGCCACCCAGCCCTCAGGTGGCA 4200  
Qy 4201 CATCTGAGGAGGAGGCTTGAAGGGCCACACACATGCTTCTTCTGGGTGAAGCCCT 4260  
Db 4201 CATCTGAGGAGGAGGCTTGAAGGGCCACACACATGCTTCTTCTGGGTGAAGCCCT 4260  
Qy 4261 TTGGCTCCGCCACTCTCTCTTGGATGGGTGTGCTCCCTTATCCCCAAATCACTCTATACA 4320  
Db 4261 TTGGCTCCGCCACTCTCTCTTGGATGGGTGTGCTCCCTTATCCCCAAATCACTCTATACA 4320  
Qy 4321 TCCAAATTCAGAAACAAACATGCTGGCAATTTCTACA CAAAAGAGATGAGATTACAGTG 4380  
Db 4321 TCCAAATTCAGAAACAAACATGCTGGCAATTTCTACA CAAAAGAGATGAGATTACAGTG 4380  
Qy 4381 CAGGTTGGGCTCTGCAATGGAGGTGCCCTATAACACAGAGAGAAAACTCTGAAAGCAC 4440  
Db 4381 CAGGTTGGGCTCTGCAATGGAGGTGCCCTATAACACAGAGAGAAAACTCTGAAAGCAC 4440  
Qy 4441 AGGGCAGGGACAGACACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAAA 4500  
Db 4441 AGGGCAGGGACAGACACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAAA 4500  
Qy 4501 ACCGAGCTGAGCATCAGGACCTTGGCTCGAATTTGTCTTCCAGTATTAGGTGCTCTTC 4560  
Db 4501 ACCGAGCTGAGCATCAGGACCTTGGCTCGAATTTGTCTTCCAGTATTAGGTGCTCTTC 4560  
Qy 4561 TCTGCCCTCTTCCACGGTATCTGTGGTTGCCAGGCTGGGAGGGAACCATAGCCAC 4620  
Db 4561 TCTGCCCTCTTCCACGGTATCTGTGGTTGCCAGGCTGGGAGGGAACCATAGCCAC 4620  
Qy 4621 ACCACAGATTTCTGAAAGTTTACAATGCAATGAGTCTTGGGGTGTAGGTGGCAGCT 4680  
Db 4621 ACCACAGATTTCTGAAAGTTTACAATGCAATGAGTCTTGGGGTGTAGGTGGCAGCT 4680  
Qy 4681 CCCCAAGGCTTGGCCCGCCAGCCCAACCATCATGACTCTAAGTGTGTATTAAATAT 4740  
Db 4681 CCCCAAGGCTTGGCCCGCCAGCCCAACCATCATGACTCTAAGTGTGTATTAAATAT 4740  
Qy 4741 TTATTTATTGGAGATGTTATTTATTAGATGATATTTATTGCAGAAATTTCTATTCTTGA 4800  
Db 4741 TTATTTATTGGAGATGTTATTTATTAGATGATATTTATTGCAGAAATTTCTATTCTTGA 4800  
Qy 4801 TTAAACAAATAAATCTTGGCCCGAGAACTTAGTCTCTTTGGCCAGCCTCACCCCTCTGG 4860  
Db 4801 TTAAACAAATAAATCTTGGCCCGAGAACTTAGTCTCTTTGGCCAGCCTCACCCCTCTGG 4860  
Qy 4861 TGCTCATCAGACTCTTGGCCAGCCCTGGCTCCCACTCCCTGCTTGGCTCTGGTGGAGCTC 4920  
Db 4861 TGCTCATCAGACTCTTGGCCAGCCCTGGCTCCCACTCCCTGCTTGGCTCTGGTGGAGCTC 4920  
Qy 4921 ACAGAGCTCTGGGAAGAGGCCCTCTTCTCTCCCGCAGCTGGGGCGATGGCGCACCTCAGA 4980  
Db 4921 ACAGAGCTCTGGGAAGAGGCCCTCTTCTCTCCCGCAGCTGGGGCGATGGCGCACCTCAGA 4980  
Qy 4981 CTTTACCACCTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCTCCACACAGCTTC 5040  
Db 4981 CTTTACCACCTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCTCCACACAGCTTC 5040  
Qy 5041 TGTCACCCAGAGTTTCCCTCACCCAGCTTTGCTAAGTCTTCTCA 5087  
Db 5041 TGTCACCCAGAGTTTCCCTCACCCAGCTTTGCTAAGTCTTCTCA 5087

## RESULT 3

AF176912

LOCUS

DEFINITION

Homo sapiens neurotrophin-1/B-cell stimulating factor-3 gene,

complete cds.

ACCESSION

AF176912

VERSION

AF176912.1

KEYWORDS

Homo sapiens.

5087 bp DNA linear

AF176912

Homo sapiens neurotrophin-1/B-cell stimulating factor-3 gene,

complete cds.

AF176912

AF176912.1

GI:6007642

Homo sapiens.

PRI 04-OCT-1999



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 5087)  
 AUTHORS Senaldi, G., Varnum, B.C., Sarmiento, U., Starnes, C., Lile, J.,  
 Scully, S., Guo, J., Elliott, G., McNinch, J., Shaklee, C.L.,  
 Freeman, D., Mand, F., Simonet, W.S., Boone, T. and Chang, M.S.  
 TITLE Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the  
 IL-6 family  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)  
 MEDLINE 99432254  
 PUBMED 10500198  
 REFERENCE 2 (bases 1 to 5087)  
 AUTHORS Senaldi, G., Varnum, B., Sarmiento, U., Lile, J., Starnes, C.,  
 Scully, S., Guo, J., Elliott, G., McNinch, J., Freeman, D., Shaklee, C.,  
 Manu, F., Simonet, S., Boone, T. and Chang, M.-S.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-AUG-1999) Angen, Inc., One Angen Center Drive,  
 Thousand Oaks, CA 91320, USA  
 FEATURES  
 source  
 1. 5087  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 join(<120..135,1368..1534,3365..>3859)  
 /product="neurotrophin-1/B-cell stimulating factor-3"  
 join(120..135,1368..1534,3365..3859)  
 /codon\_start=1  
 /product="neurotrophin-1/B-cell stimulating factor-3"  
 /protein\_id="AAFO0992.1"  
 /db\_xref="GI:6007643"  
 /translation="MDLRAGDSWGLACLCTVLWHLPAVPAALNRTGDPGSPISQKTY  
 DLTRYLHQJRSAGTYLYNIPFPNPRLLAGETLPRATVDLEWVRSLNDKLR  
 LTQNEYASHLLCYRLGNLQAAFAELRLAHFCTSLQGLGSLAGVMAALGVPLPQ  
 PLPGTEPTLPQPAHSDFLQKMDDEWLLKELQTLWLSAKDFENRLKKMQPPAAAVTL  
 HLGAHGP"  
 BASE COUNT 992 a 1746 c 1191 g 1158 t  
 ORIGIN  
 Query Match 100.0%; Score 5087; DB 9; Length 5087;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AACCTGGAGTGGGCTGGCGATGGGATTTAAAGCTTCGCGAGCGCGGCTGCC 60  
 DB 1 AACCTGGAGTGGGCTGGCGATGGGATTTAAAGCTTCGCGAGCGCGGCTGCC 60  
 QY 61 CTCCCACTCCCGAGCTCCGGAGAGGAGCGGACCGCGCGCGCGCGCGCGCCCA 120  
 DB 61 CTCCCACTCCCGAGCTCCGGAGAGGAGCGGACCGCGCGCGCGCGCGCGCCCA 120  
 QY 121 TGGACCTCCGAGCGTTGAAAACCCAACTAGCCCTCTCTTATACATGACAAAGCAG 180  
 DB 121 TGGACCTCCGAGCGTTGAAAACCCAACTAGCCCTCTCTTATACATGACAAAGCAG 180  
 QY 181 CGCCCATCTGATCTAAACCGACCAAGTCACAGCTCCCACTACCCCTTCGCTGCC 240  
 DB 181 CGCCCATCTGATCTAAACCGACCAAGTCACAGCTCCCACTACCCCTTCGCTGCC 240  
 QY 241 CAGACCTCACCACATCTTGGAGTCAAACTCAACCGCTAAATCAACCAATCCCA 300  
 DB 241 CAGACCTCACCACATCTTGGAGTCAAACTCAACCGCTAAATCAACCAATCCCA 300  
 QY 301 AGTCTAAACTAATCTGAACTTTTAAAGTAACCCAGTCTTTAAACCTAGCCCAAT 360  
 DB 301 AGTCTAAACTAATCTGAACTTTTAAAGTAACCCAGTCTTTAAACCTAGCCCAAT 360  
 QY 361 GCCAATATATCTACCTAGCCAAACCTAATGCTCTTGGCAGTCCAAAGTGTCCATG 420  
 DB 361 GCCAATATATCTACCTAGCCAAACCTAATGCTCTTGGCAGTCCAAAGTGTCCATG 420  
 QY 421 AATCCTCACCTTGTCTCACTGAAATCCAGAAAGCATATTTCCCACTGCCACAT 480  
 DB 421 AATCCTCACCTTGTCTCACTGAAATCCAGAAAGCATATTTCCCACTGCCACAT 480

481 CCCTCTTTACAGACCCCAACCCCTGGCTCTGACTCTCTGGTATCTCTGGATGTCCAAACT 540  
 DB 481 CCCTCTTTACAGACCCCAACCCCTGGCTCTGACTCTCTGGTATCTCTGGATGTCCAAACT 540  
 QY 541 CTGAGTGCATTCAGCCCAACAGCCCGAGCTCGTAAATGACCTCTCTCCCTTCCTGTC 600  
 DB 541 CTGAGTGCATTCAGCCCAACAGCCCGAGCTCGTAAATGACCTCTCTCCCTTCCTGTC 600  
 QY 601 CCACCTTTCAGGCTGTATGGAAGAGCCCTCATTTGAAGTCCAACTTTTCCCACTAACACC 660  
 DB 601 CCACCTTTCAGGCTGTATGGAAGAGCCCTCATTTGAAGTCCAACTTTTCCCACTAACACC 660  
 QY 661 AAGAAACGGGGTGAACCTCCACACTGCCACCGCTTCCTCGAGAGTGAGCACTAAATCTCCTT 720  
 DB 661 AAGAAACGGGGTGAACCTCCACACTGCCACCGCTTCCTCGAGAGTGAGCACTAAATCTCCTT 720  
 QY 721 CAATCTAACCCACCTTACACTTCCCACTCAGAGAAATCAATCTAGAAATATACCCAAA 780  
 DB 721 CAATCTAACCCACCTTACACTTCCCACTCAGAGAAATCAATCTAGAAATATACCCAAA 780  
 QY 781 ACTAAGCCCAATAGGAGAGCCGACCTAGTGTCTAACCCCTATACCTTGTCTCTATGG 840  
 DB 781 ACTAAGCCCAATAGGAGAGCCGACCTAGTGTCTAACCCCTATACCTTGTCTCTATGG 840  
 QY 841 GTGAGTCTGTCTTGGGGCGCGCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
 DB 841 GTGAGTCTGTCTTGGGGCGCGCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
 QY 901 AGCTGCGAGCTCTGACATGTGTCTCCACCTCTGACTCTGCTCTCTCTCTCTCTCTCTCTCT 960  
 DB 901 AGCTGCGAGCTCTGACATGTGTCTCCACCTCTGACTCTGCTCTCTCTCTCTCTCTCTCTCT 960  
 QY 961 GACTGGAGACTGCGAGGAGCTAGGTACAACTGGGACACAGGAGCTGCACTGCACT 1020  
 DB 961 GACTGGAGACTGCGAGGAGCTAGGTACAACTGGGACACAGGAGCTGCACTGCACT 1020  
 QY 1021 CCCTAGGCTGGCCCGCTCCCTCCATGTACACATATATACATGTGTGGACACACACAGTG 1080  
 DB 1021 CCCTAGGCTGGCCCGCTCCCTCCATGTACACATATATACATGTGTGGACACACACAGTG 1080  
 QY 1081 GCACATATGCAAAAGACTCTCTCAGCTGACACACAGATCCATTTCTCAAGTATCTACTGAT 1140  
 DB 1081 GCACATATGCAAAAGACTCTCTCAGCTGACACACAGATCCATTTCTCAAGTATCTACTGAT 1140  
 QY 1141 AGACACTCATGCTGCGCAAGTCTCTCATCTCAAAACATACATGCTCTCTCTCTCTCTCT 1200  
 DB 1141 AGACACTCATGCTGCGCAAGTCTCTCATCTCAAAACATACATGCTCTCTCTCTCTCTCT 1200  
 QY 1201 GTCTTGGCAGGAGTGTCTTCCCTCTCTCCATCCCTCTGCTCCCTCCATCTGGTGTCCACCC 1260  
 DB 1201 GTCTTGGCAGGAGTGTCTTCCCTCTCTCCATCCCTCTGCTCCCTCCATCTGGTGTCCACCC 1260  
 QY 1261 TCACCCCCCAAGCTGCGGAGACACACCTGAGGGGCTGCCAGCTGCTTCCCTCC 1320  
 DB 1261 TCACCCCCCAAGCTGCGGAGACACACCTGAGGGGCTGCCAGCTGCTTCCCTCC 1320  
 QY 1321 GTGTGGGCGCGGCGGCTCATGCTTCTGCTCCATCTGCTCCCAAGGGGACTCGTGGG 1380  
 DB 1321 GTGTGGGCGCGGCGGCTCATGCTTCTGCTCCATCTGCTCCCAAGGGGACTCGTGGG 1380  
 QY 1381 GATGTAGCGTCTGTGCAAGTGTCTGGCACTCTCCCTGAGTCCAGCTCTCAATCG 1440  
 DB 1381 GATGTAGCGTCTGTGCAAGTGTCTGGCACTCTCCCTGAGTCCAGCTCTCAATCG 1440  
 QY 1441 CACAGGGGACCCAGGGCTGCGCCCTCCATCCAGAAACCTTATGACCTCACCCGCTACCT 1500  
 DB 1441 CACAGGGGACCCAGGGCTGCGCCCTCCATCCAGAAACCTTATGACCTCACCCGCTACCT 1500  
 QY 1501 GGAGCACCAACTCGCGAGCTTGGCTGGGACCTATGTAGTATCCAGCGTAGGAATCTGGG 1560  
 DB 1501 GGAGCACCAACTCGCGAGCTTGGCTGGGACCTATGTAGTATCCAGCGTAGGAATCTGGG 1560















Db 144122 AATGATATCCCTTATTTGAAAAAAGCTCAGGGAGGCCATGAAACAAAGAGCCCTAGCATGGAGA 144063  
Qy 3252 CAGGCGCAGTGTGAGGGGACACAAAAATAGAAACTTTTGGGAGCAGGTATCTCTTGGTG 3311  
Db 144062 CAGGCGCAGTGTGAGGGGACACAAAAATAGAAACTTTTGGGAGCAGGTATCTCTTGGTG 144003  
Qy 3312 GTGAGCAGCGGCTGCGCCCTCTCTCTCCCATCAGCCCTCTCTCTTTTACAGCTGAAC 3371  
Db 144002 GTGAGCAGCGGCTGCGCCCTCTCTCTCCCATCAGCCCTCTCTCTTTTACAGCTGAAC 143943  
Qy 3372 ACCTGGGCCCCCTTTCAACAGCAGCAGACTTCAACCCCTCCCGCTGGGGGCGAGAGCTC 3431  
Db 143942 ACCTGGGCCCCCTTTCAACAGCAGCAGACTTCAACCCCTCCCGCTGGGGGCGAGAGCTC 143883  
Qy 3432 TGCCCAAGGCGCACTGTGTGACTTGGAGGTGAGGAGAGCCCTCAATGACAAACTGGGGCTGA 3491  
Db 143882 TGCCCAAGGCGCACTGTGTGACTTGGAGGTGAGGAGAGCCCTCAATGACAAACTGGGGCTGA 143823  
Qy 3492 CCAGAACTACGAGGCGCTACAGCCACCTCTGTGTGTTACTTTCGCTGGGCTCAACCGCTCAGG 3551  
Db 143822 CCAGAACTACGAGGCGCTACAGCCACCTCTGTGTGTTACTTTCGCTGGGCTCAACCGCTCAGG 143763  
Qy 3552 CTGCCACTGTGAGCTGCGCGCAGCCTGCGCCACTTCTGACACAGCCTCCAGGCGCTGC 3611  
Db 143762 CTGCCACTGTGAGCTGCGCGCAGCCTGCGCCACTTCTGACACAGCCTCCAGGCGCTGC 143703  
Qy 3612 TGGGCGAGCATGCGGGGCTCATGGCAGCTCTGGGCTACCCACTGCCCCAGCGCTGCTG 3671  
Db 143702 TGGGCGAGCATGCGGGGCTCATGGCAGCTCTGGGCTACCCACTGCCCCAGCGCTGCTG 143643  
Qy 3672 GGAAGTGAACCACTTGAGCTCTGGGCTGCGCCCTGCGCAGTGACTTCTCCAGAAAGATGGAAC 3731  
Db 143642 GGAAGTGAACCACTTGAGCTCTGGGCTGCGCCCTGCGCAGTGACTTCTCCAGAAAGATGGAAC 143583  
Qy 3732 ACTTCTGGCTGCTGAAGAGAGCTGAGACCTGCGCTGCTGCGCTGCGGCGCAAGACTTCAAC 3791  
Db 143582 ACTTCTGGCTGCTGAAGAGAGCTGAGACCTGCGCTGCTGCGCTGCGGCGCAAGACTTCAAC 143523  
Qy 3792 GGCTCAAGAAAGATGAGGCTCCAGGCTCCAGAGCTGAGTCAACCTGACCTGGGGGCTCATG 3851  
Db 143522 GGCTCAAGAAAGATGAGGCTCCAGGCTCCAGAGCTGAGTCAACCTGACCTGGGGGCTCATG 143463  
Qy 3852 GCTTCTGACTTCTGACCTTCTCTCTCTGCTCCCTTCAAACTGCTGCTCCACTTTGTG 3911  
Db 143462 GCTTCTGACTTCTGACCTTCTCTCTCTGCTCCCTTCAAACTGCTGCTCCACTTTGTG 143403  
Qy 3912 AGAGCCAGCCCTGTATGCAACACCTGTGTGAGCCAGAGAGAGAGCTGTGAGCCTCTGG 3971  
Db 143402 AGAGCCAGCCCTGTATGCAACACCTGTGTGAGCCAGAGAGAGAGCTGTGAGCCTCTGG 143343  
Qy 3972 CCCTTCTGAGCCGCTGGGCGTGTGATGCGATCAGCCCTGTCTCTCCCTCCCACTCCCA 4031  
Db 143342 CCCTTCTGAGCCGCTGGGCGTGTGATGCGATCAGCCCTGTCTCTCCCTCCCACTCCCA 143283  
Qy 4032 AAGGTCTACCGAGCTGGGGAGGAGTACAGTAGGCGCTGTCTCTGCTCTGTTTCTACAGGA 4091  
Db 143282 AAGGTCTACCGAGCTGGGGAGGAGTACAGTAGGCGCTGTCTCTGCTCTGTTTCTACAGGA 143223  
Qy 4092 AGTCATGCTCAGGAGGTGTGAAGTGTGTGAGTGTGTGAGGAGGCTCATGGGCTCTCT 4151  
Db 143222 AGTCATGCTCAGGAGGTGTGAAGTGTGTGAGTGTGTGAGGAGGCTCATGGGCTCTCT 143163  
Qy 4152 GCTTCTTCCCTACCACTTGGGCGAGTCCACCCAGCCCTCAGGTGGGACATCTGGAGGG 4211  
Db 143162 GCTTCTTCCCTACCACTTGGGCGAGTCCACCCAGCCCTCAGGTGGGACATCTGGAGGG 143103  
Qy 4212 CAGGGGTTGAGGGGCCACACACATGCTCTTCTGGGGTGAAGCCCTTTGGCTGCCCTC 4271  
Db 143102 CAGGGGTTGAGGGGCCACACACATGCTCTTCTGGGGTGAAGCCCTTTGGCTGCCCTC 143043  
Qy 4272 ACTCTCTTGGATGGGTGTGCTCTTCTTCCCAATCACTCTATACATCCAAATTCAGG 4331  
Db 143042 ACTCTCTTGGATGGGTGTGCTCTTCTTCCCAATCACTCTATACATCCAAATTCAGG 142983

Qy 4332 AAACAAACATGGTGGCAATTTCTACAAAAAGAGATGAGATTAAACAGTGCAGGGTTGGGG 4391  
Db 142982 AAACAAACATGGTGGCAATTTCTACAAAAAGAGATGAGATTAAACAGTGCAGGGTTGGGG 142923  
Qy 4392 TCTGCAATGGAGGTGCCCTATAAACACAGAGAGAAATACCTGAAAGCAGACAGGGGAGGGA 4451  
Db 142922 TCTGCAATGGAGGTGCCCTATAAACACAGAGAGAAATACCTGAAAGCAGACAGGGGAGGGA 142863  
Qy 4452 CAGACACAGACAGACCCAGGAGTCTCCAAAGCAGAGCTGGCAAAACAAACCCAGAGCTGA 4511  
Db 142862 CAGACACAGACAGACCCAGGAGTCTCCAAAGCAGAGCTGGCAAAACAAACCCAGAGCTGA 142803  
Qy 4512 GCATCAGGACCTTGCCTCGAAATTTGTTTCCAGTATTACGGTGCCTTCTCTGCCCCCTT 4571  
Db 142802 GCATCAGGACCTTGCCTCGAAATTTGTTTCCAGTATTACGGTGCCTTCTCTGCCCCCTT 142743  
Qy 4572 TCCAGAGGTATCTGTGGTGTCCAGGCTGGGAGGCGCAACCATAGCCACACACAGAGTT 4631  
Db 142742 TCCAGAGGTATCTGTGGTGTCCAGGCTGGGAGGCGCAACCATAGCCACACACAGAGTT 142683  
Qy 4632 TCCTGAAAGTTTACAAATGCAGTAGCATTTTGGGGTGTAGGGTGGCAGCTCCCCAAGGCC 4691  
Db 142682 TCCTGAAAGTTTACAAATGCAGTAGCATTTTGGGGTGTAGGGTGGCAGCTCCCCAAGGCC 142623  
Qy 4692 TGCCCCCAGGCCCCACCACTCATGACTTAAGTGTGTGTTGTTATTAATTTATTATTG 4751  
Db 142622 TGCCCCCAGGCCCCACCACTCATGACTTAAGTGTGTGTTGTTATTAATTTATTATTG 142563  
Qy 4752 GAGATGTTATTATTAGATGATTTATTGACAGATTTCTATTCTTGTTATTAACAATAA 4811  
Db 142562 GAGATGTTATTATTAGATGATTTATTGACAGATTTCTATTCTTGTTATTAACAATAA 142503  
Qy 4812 AATGCTGCCCGCAGAACTTAGTCTCTTTGCCCGAGCTCACCCCTCTGCTGCTCATCAGA 4871  
Db 142502 AATGCTGCCCGCAGAACTTAGTCTCTTTGCCCGAGCTCACCCCTCTGCTGCTCATCAGA 142443  
Qy 4872 CTCTTGCCACCCCTGGTCCCACTCCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4931  
Db 142442 CTCTTGCCACCCCTGGTCCCACTCCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 142383  
Qy 4932 GGAAGAGGCGCTTCTCCCGCAGTGGGCGATGGGCGCACTCAGACTTACCCACTG 4991  
Db 142382 GGAAGAGGCGCTTCTCCCGCAGTGGGCGATGGGCGCACTCAGACTTACCCACTG 142323  
Qy 4992 CTGCTGCCACCAACCCCTTGATCCCTCAGTCTCCCAACAGCTTCTGTCACCCCA 5051  
Db 142322 CTGCTGCCACCAACCCCTTGATCCCTCAGTCTCCCAACAGCTTCTGTCACCCCA 142263  
Qy 5052 GGTTCCTCTCACCACCCCTTGTGCTAAGTCTTCTCA 5087  
Db 142262 GGTTCCTCTCACCACCCCTTGTGCTAAGTCTTCTCA 142227

RESULT 6  
AP002437 135116 bp DNA linear HTG 06-JUN-2000  
LOCUS Homo sapiens chromosome 11 clone RP11-678D20 map 11q13, WORKING  
DEFINITION DRAFT SEQUENCE, 41 unordered pieces.  
ACCESSION AP002437  
VERSION AP002437.1 GI:8307741  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 135116)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Iwata,T., Tokoka,Y., Watanabe,H. and Sakaki,Y.  
TITLE Homo sapiens 135,116 genomic DNA of 11q13  
JOURNAL Published Only in DataBase (2000)  
REFERENCE 2 (bases 1 to 135116)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,



Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
 Direct Submission  
 Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,  
 Japan (E-mail: hattori@gsc.riken.go.jp,  
 URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,  
 Fax: 81-42-778-9924)

## COMMENT

----- Genome Center  
 Center: RIKEN Genomic Sciences Center (GSC)  
 Center code: RIKEN  
 Web site: http://hgp.gsc.riken.go.jp/  
 Contact: hattori@gsc.riken.go.jp  
 ----- Project Information  
 Center project name: HumDraft11  
 Center clone name: RP11-678D20  
 ----- Summary Statistics  
 Sequencing vector: PCR products; 100% of reads  
 Chemistry: Dye-terminator ET-amersham; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 109313 bases at least Q40  
 Consensus quality: 121923 bases at least Q30  
 Consensus quality: 127981 bases at least Q20  
 Insert size: 131116; sum-of-contigs  
 Quality coverage: 4.18x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of  
 41 contigs. The true order of the pieces is not known and their  
 order in this sequence record is arbitrary. Gaps between the  
 contigs are represented as runs N, but the exact sizes of the gaps  
 are unknown. This record will be updated with the finished sequence  
 as soon as it is available and the accession number will be  
 preserved

1 7749 contig of 7749 bp in length  
 7850 15108 contig of 7259 bp in length  
 15209 23728 contig of 8520 bp in length  
 23829 29302 contig of 5474 bp in length  
 29403 36171 contig of 6769 bp in length  
 36272 42123 contig of 5852 bp in length  
 42224 48172 contig of 5949 bp in length  
 48273 52700 contig of 4787 bp in length  
 52801 57588 contig of 4428 bp in length  
 57688 61624 contig of 3937 bp in length  
 61725 67756 contig of 3078 bp in length  
 67857 71834 contig of 3978 bp in length  
 71935 76800 contig of 4866 bp in length  
 76901 81069 contig of 4169 bp in length  
 81170 81932 contig of 763 bp in length  
 82033 86213 contig of 4181 bp in length  
 86314 89537 contig of 3224 bp in length  
 89638 92372 contig of 2735 bp in length  
 92473 94979 contig of 2507 bp in length  
 95080 96272 contig of 1193 bp in length  
 96373 98109 contig of 1737 bp in length  
 98210 101030 contig of 2821 bp in length  
 101131 103760 contig of 2630 bp in length  
 103861 106330 contig of 2470 bp in length  
 106431 108601 contig of 2171 bp in length  
 108702 110440 contig of 1739 bp in length  
 110541 112280 contig of 2033 bp in length  
 112381 114885 contig of 1740 bp in length  
 114986 117018 contig of 2029 bp in length  
 117119 118839 contig of 1721 bp in length  
 118940 120968 contig of 1062 bp in length  
 121069 123048 contig of 1879 bp in length  
 123048 124231 contig of 1396 bp in length  
 124332 125727 contig of 1365 bp in length  
 125828 127192 contig of 1727 bp in length  
 127293 129019 contig of 1626 bp in length  
 129120 130181 contig of 1202 bp in length  
 130282 131483 contig of 1012 bp in length  
 131584 132595 contig of 1191 bp in length  
 132696 133886 contig of

133987 135116 contig of 1130 bp in length.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 41 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 7749 contig of 7749 bp in length  
 7750 7849 gap of 100 bp  
 7850 15108 contig of 7259 bp in length  
 15109 15208 gap of 100 bp  
 15209 23728 contig of 8520 bp in length  
 23729 23828 gap of 100 bp  
 23829 29302 contig of 5474 bp in length  
 29303 29402 gap of 100 bp  
 29403 36171 contig of 6769 bp in length  
 36172 36271 gap of 100 bp  
 36272 42123 contig of 5852 bp in length  
 42124 42223 gap of 100 bp  
 42224 48172 contig of 5949 bp in length  
 48173 48272 gap of 100 bp  
 48273 52700 contig of 4428 bp in length  
 52701 52800 gap of 100 bp  
 52801 57587 contig of 4787 bp in length  
 57588 57687 gap of 100 bp  
 57688 61624 contig of 3937 bp in length  
 61625 61724 gap of 100 bp  
 61725 67756 contig of 6032 bp in length  
 67757 67856 gap of 100 bp  
 67857 71834 contig of 3978 bp in length  
 71835 71934 gap of 100 bp  
 71935 76800 contig of 4866 bp in length  
 76801 76900 gap of 100 bp  
 76901 81069 contig of 4169 bp in length  
 81070 81169 gap of 100 bp  
 81170 81932 contig of 763 bp in length  
 81933 82032 gap of 100 bp  
 82033 86213 contig of 4181 bp in length  
 86214 86313 gap of 100 bp  
 86314 89537 contig of 3224 bp in length  
 89538 89637 gap of 100 bp  
 89638 92372 contig of 2735 bp in length  
 92373 92472 gap of 100 bp  
 92473 94979 contig of 2507 bp in length  
 94980 95079 gap of 100 bp  
 95080 96272 contig of 1193 bp in length  
 96273 96372 gap of 100 bp  
 96373 98109 contig of 1737 bp in length  
 98110 98209 gap of 100 bp  
 98210 101030 contig of 2821 bp in length  
 101031 101130 gap of 100 bp  
 101131 103760 contig of 2630 bp in length  
 103761 103860 gap of 100 bp  
 103861 106330 contig of 2470 bp in length  
 106331 106430 gap of 100 bp  
 106431 108601 contig of 2171 bp in length  
 108602 108701 gap of 100 bp  
 108702 110440 contig of 1739 bp in length  
 110441 110540 gap of 100 bp  
 110541 112280 contig of 1740 bp in length  
 112281 112380 gap of 100 bp  
 112381 114885 contig of 2029 bp in length  
 114886 114985 gap of 100 bp  
 114986 117018 contig of 2033 bp in length  
 117019 117118 gap of 100 bp  
 117119 118839 contig of 1721 bp in length  
 118840 118939 gap of 100 bp  
 118940 120968 contig of 2029 bp in length  
 120969 121068 gap of 100 bp  
 121069 122947 contig of 1879 bp in length  
 122948 123047 gap of 100 bp







Assembly program: Phrap; version 0.960731  
 Consensus quality: 267668 bases at least Q40  
 Consensus quality: 268459 bases at least Q30  
 Consensus quality: 268627 bases at least Q20  
 Insert size: 242000; agarose-fp  
 Insert size: 268755; sum-of-contigs  
 Quality coverage: 10.4 in Q20 bases; agarose-fp  
 Quality coverage: 9.4 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 100642: contig of 100642 bp in length  
 \* 100643 100742: gap of 100 bp  
 \* 100743 104988: contig of 4246 bp in length  
 \* 104989 105088: gap of 100 bp  
 \* 105089 112582: contig of 7494 bp in length  
 \* 112583 112682: gap of 100 bp  
 \* 112683 142041: contig of 29359 bp in length  
 \* 142042 142141: gap of 100 bp  
 \* 142142 179941: contig of 37800 bp in length  
 \* 179942 180041: gap of 100 bp  
 \* 180042 259317: contig of 79276 bp in length  
 \* 259318 259417: gap of 100 bp  
 \* 259418 269355: contig of 9938 bp in length.

#### FEATURES

Location/Qualifiers  
 1..269355  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-41B18"  
 /clone\_lib="RPCI-23 Female Mouse BAC"  
 misc\_feature  
 1..105642  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 misc\_feature  
 100743..104988  
 /note="assembly\_fragment"  
 misc\_feature  
 105089..112582  
 /note="assembly\_fragment"  
 misc\_feature  
 112683..142041  
 /note="assembly\_fragment"  
 misc\_feature  
 142142..179941  
 /note="assembly\_fragment"  
 misc\_feature  
 180042..259317  
 /note="assembly\_fragment"  
 misc\_feature  
 259418..269355  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:right

BASE COUNT 67479 a 67954 c 68125 g 65191 t 606 others  
 ORIGIN

Query Match 33.6%; Score 1709.4; DB 2; Length 269355;  
 Best Local Similarity 72.6%; Pred. No. 0;  
 Matches 2887; Conservative 0; Mismatches 861; Indels 231; Gaps 43;

QY 898 CTAGCTGCCAGCTCTGACATGCTGTCTCCACCTCTGACTCCCTCAAGCTGAG 957  
 Db 148985 CACAGCTTGGCAGCTGACATATGCCACATTCACCCCTTAATTC-----TGATTCA 149039  
 QY 958 TGGACTGGAAGACTGCGAG--GAAGCTAGGGTAACTGGAACACACAGGAGGTCGACCT 1015  
 Db 149040 TGGACTGGAAGGCTTGGAGAGAGACTGGGACACACTGAAGCTTAGGCATATTGAC 149099  
 QY 1016 GCAGTCCCTAGGCTGGCCCCGCTCCATGTACACATATACATATGTTGGCACACACA 1075  
 Db 149100 AAAATCCCTAAAGCTTGGCTGCCCTTTGAATAGACACATGACAC----- 149145

QY 1076 CAGTGGCACACATGCCAAAGACTCTCTCAGCTGACACACACAGATCCATCTTCAAGTATCTA 1135  
 Db 149146 -----CATATTTGGCAGATACATGCAAGAGACTTTTCAGCTAATAGCAGTCAAGTATCTA 149199  
 QY 1136 CTGATAGACACTCATGCTGCCAAGCTCTCATCTCTCAAAACATACACATGCTCTCTTCT 1195  
 Db 149200 CTGATAGACACAATGCCAGATTGTCT-----TTGCATACATACCTCCACAT 149248  
 QY 1196 CTCCCTCTTGGCAGAGTGTTCCTCTCTCCATCCCTCTGCTCCCTCCATCTGGTGTCC 1255  
 Db 149249 AGCTCTTGGACCAAGTCTTTCCCTCTCTCCACCCCTCCATTTCCCATCTGGTCTCT 149308  
 QY 1256 CACCTCTACCCCTCC-ACCCAGCCCAAGTGGGGACAGACACCTGAGGGGTGCGAGTGC 1314  
 Db 149309 CTCCCTCACCCCTCTTCTAGCCCAAGTGGGGACAGACACCTGAGGGGTGCGAGTGT 149368  
 QY 1315 TTCCCGTGTGGGCCCGGG-CCGCGCTCATGTTCTCTCTCCATCTCTCCACAGGGGACT 1373  
 Db 149369 TTCCCTATTTGGGCTTGGGCCCACTCTCATCTTCTTGGCCATCTTATTCACAGGGGACT 149428  
 QY 1374 CGTGGGGATGTAGGTCCTGTGCA CGGTCTCTTGGCACTCTCCCTGAGTGCCAGCTC 1433  
 Db 149429 CGTGGGGATGTAGCTTGCCTATGCA CGGTCTGTGGCACCTCTCCCTGAGTGCCAGCTC 149488  
 QY 1434 TCAATCGCACAGGGACCCAGGGCTGGGCCCTCCATCCAGAAACCTATGACCTCACCC 1493  
 Db 149489 TTAATCGCACAGAGATCCAGGCCCTGGGCCCTTCCATCCAGAAACCTATGACCTCACCC 149548  
 QY 1494 GCTACTGGAGCACCAACTCCGACGTTGGCTGGGACCTATGTAGTATCCAGCGTAGGA 1553  
 Db 149549 GCTACTGGAGCATCACTCCGACGTTAGCTGGGACCTAGCTAGTATTCCTTTGGCA 149608  
 QY 1554 ATCTGGGAGTTGGGAGAGAGTGGGGAAGACAGTCTTAACCGTGGAGGGTTC 1613  
 Db 149609 A-----AGAGGAGGACAGAGGCTTTGGGGAAGAGAGATGGGGCAGGGTTC 149656  
 QY 1614 TGGTAAATGATGGGTGAGGAGGGCTCTTTGGCTCCACAGTCCCTCTGTGGTCTA 1673  
 Db 149657 TAGTGAATGATGGGTGATGAGGCCCTTTGGGTCCACATAGCTTCCATCTGGCTTA 149716  
 QY 1674 TCTCTGCTCTTCCCTCTTAGTGTGGCCCCCCTCCCTGCCCCCAGGACTAG 1733  
 Db 149717 TCTTCTGCTCTTCCCTCTCAGGTGCCCT-----CCCCAATCTTGGCCCCAGGATAG 149769  
 QY 1734 GCATGTGGGACAGGCTTGCACCCCGCTTGGCCCATTTGCCCATCTGGCTGCCAGCCAGCC 1793  
 Db 149770 GCATGTGGGACAGGCTTGCACCCCGCTTGGCCCATTTGCCCATTTGGCTGCCAGCCAGCT 149829  
 QY 1794 GCCCGCTTCCCTCTGGGGCCGGGGAAGTCTCTCTGTTTACCGTCTTGGTGTCTC 1853  
 Db 149830 GCCTGCTTCCCTCTGGGGCTGGGGAGTCTCTCTGTTTACCGTCTTGGTGTCTC 149889  
 QY 1854 TTGGCGGGGGGTGGGTGGGACAGAGGGGCCCACTCCATGCTGCTGGTTCACAGC 1913  
 Db 149890 TTGAAGGGTGGGGCTGGGTGGGATGGAGGGGCCCACTTCTATGTTGTTCCAGC 149949  
 QY 1914 TCGCTCTGCCCCCAGACCTGGGGCCCTGTGCTCTGACCCAGGGGCTCTCTTCCCTC 1973  
 Db 149950 TCGC--CTACCCCCAGACCTTGGGCCCTGTGTTTCTGAGCCATGGGGCTCTCTCTC 150006  
 QY 1974 TGCTCTTCCATCTTACCTGCTGGGCTCTTAGGGGGTCTAGGGGAGAGGGGACTGTAGGA 2033  
 Db 150007 TGCTCTTCCATCTTACAGACCTCTTAGGGG-----TTGGAGAGGAGATCTACAAGAA 150063  
 QY 2034 ACCAGGACGTAGTGGCAGGGGTTTAGGTGTGGATGGAGGTTATGCTGTGAAGATTG 2093  
 Db 150064 TCTAGGACAGTCTAGGCAAGGGTTTACGGGTATGGAGGGTGTGTGCCATAAGAAATTG 150123  
 QY 2094 GGGGTGTCTAGAGTGTCTCAGAGAGCCAGGAGAGAGAGAGAGGGTGTGGAGGAGCCG 2153  
 Db 150124 AGGGTAGTTTAGAGATGTTTACAGAGCCAGGAGAGAGAGAGAGGATCGGAGGACCTG 150183







3963 AGCTCTGGCCCTTCTCTGGACCGGCTGGCGTGTGATCGGATCAGCCCTGTCTCTCC 4022  
Db 854 AGCTCTGGCCCTTCTCTGGACCGGCTGGCGTGTGATCGGATCAGCCCTGTCTCTCC 913  
Qy 4023 CACCTCCCAAGGCTTACCGAGCTGGGAGGAGTACAGTAGGCGCTGTCTCTCTGTT 4082  
Db 914 CACCTCCCAAGGCTTACCGAGCTGGGAGGAGTACAGTAGGCGCTGTCTCTCTGTT 973  
Qy 4083 TCTACAGGAAGTCATGCTCGAGGAGTGTGAAGTGGTTCAGGTGTGGTGCAGAGCGCTCA 4142  
Db 974 TCTACAGGAAGTCATGCTCGAGGAGTGTGAAGTGGTTCAGGTGTGGTGCAGAGCGCTCA 1033  
Qy 4143 TGGCT 4202  
Db 1034 TGGCT 1093  
Qy 4203 TCTGAGGGGAGGGGTTGAGGGGCGCCACCCACACATGCTCTCTCTCTCTCTCTCTCT 4262  
Db 1094 TCTGAGGGGAGGGGTTGAGGGGCGCCACCCACACATGCTCTCTCTCTCTCTCTCTCT 1153  
Qy 4263 GGCTGCCCACT 4322  
Db 1154 GGCTGCCCACT 1213  
Qy 4323 CAATTGAGGAACAATGCTGGCAATTTCTACACAAAAGAGATGAGATTAACAGTGCA 4382  
Db 1214 CAATTGAGGAACAATGCTGGCAATTTCTACACAAAAGAGATGAGATTAACAGTGCA 1273  
Qy 4383 GGGTGGGGTCTGCATTTGAGGTTGCCCTATAAAACAGAGAGAAATACATGAAAGCACAG 4442  
Db 1274 GGGTGGGGTCTGCATTTGAGGTTGCCCTATAAAACAGAGAGAAATACATGAAAGCACAG 1333  
Qy 4443 GGCAGGAGCAGACACAGACAGAGAGAGTCTCCAAAGCAGAGAGTGGCAACCAAC 4502  
Db 1334 GGCAGGAGCAGACACAGACAGAGAGTCTCCAAAGCAGAGAGTGGCAACCAAC 1393  
Qy 4503 CCGAGCTGAGCATCAGAGACCTTGCCTCGAATTTCTCCAGTATTAACGGTGGCTCTCTCT 4562  
Db 1394 CCGAGCTGAGCATCAGAGACCTTGCCTCGAATTTCTCCAGTATTAACGGTGGCTCTCTCT 1453  
Qy 4563 TGCCCCCTTTCCAGAGGTATCTGTGGTGTCCAGGCTGGGAGGCAACCATAGGCACAC 4622  
Db 1454 TGCCCCCTTTCCAGAGGTATCTGTGGTGTCCAGGCTGGGAGGCAACCATAGGCACAC 1513  
Qy 4623 CACAGGATTTCTGAAAGTTTACATGACAGTAGCAATTTTGGGTGTAGGTGGCAGCTCC 4682  
Db 1514 CACAGGATTTCTGAAAGTTTACATGACAGTAGCAATTTTGGGTGTAGGTGGCAGCTCC 1573  
Qy 4683 CCAAGGCTTCCGCCCCCAGCCCCCAGCTCATGACTCTAAGTGTGTGTAATTAATTT 4742  
Db 1574 CCAAGGCTTCCGCCCCCAGCCCCCAGCTCATGACTCTAAGTGTGTGTAATTAATTT 1633  
Qy 4743 ATTTATTGGAGATGTTATTATTATTAGATGATTTATTATTGAGAAATTTCTATTCTGATT 4802  
Db 1634 ATTTATTGGAGATGTTATTATTATTAGATGATTTATTATTGAGAAATTTCTATTCTGATT 1693  
Qy 4803 AACAAATAAATGCTTGGCCCGAGAA 4827  
Db 1694 AACAAATAAATGCTTGGCCCGAGAA 1718  
RESULT 9  
AF172854 AF172854 1689 bp mRNA linear PRI 10-SEP-1999  
LOCUS  
DEFINITION Homo sapiens cardiotrophin-like cytokine CLC mRNA, complete cds.  
ACCESSION AF172854  
VERSION AF172854.1 GI:5952980  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1. (bases 1 to 1689)

AUTHORS Shi, Y., Wang, W., Yourey, P.A., Gohari, S., Zukauskas, D., Zhang, J., Ruben, S. and Alderson, R.F.  
TITLE Computational EST database analysis identifies a novel member of the neuropoietic cytokine family  
JOURNAL Biochem. Biophys. Res. Commun. 262 (1), 132-138 (1999)  
MEDLINE 99382254  
PUBMED 1048081  
REFERENCE 2 (bases 1 to 1689)  
AUTHORS Shi, Y.  
TITLE Submitted (28-JUL-1999) Molecular Biology, Human Genome Science, Inc., 9410 Key West Avenue, Rockville, MD 20850, USA  
JOURNAL Location/Qualifiers  
FEATURES  
source  
1. .1689  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11q13.3"  
46. .723  
/codon\_start=1  
/product="cardiotrophin-like cytokine CLC"  
/protein\_id="AAD54284.1"  
/db\_xref="GI:5852981"  
/translation="MDLRAGDSWGLAELCTVLMHLPVAPALNRGDPGPGPSIOIKTYDLTRYLEHQRLSLAGTLYNLYLGPFPNEDFPNRLGAETLPRATVDLEWVRSLNDKLRLTQNYEAYSHLLCYLRGNRQAATAELELRSLAHFCTSLQGLLSGSIAGVMAALGYPLPQPLPGTEPTWTTPGPAHSDFLQKMDDFLLKELQTLWRSKDFNRLKKMQPPAAAVTLHLGAHGF"  
BASE COUNT 349 a 530 c 448 g 362 t  
ORIGIN  
Query Match 28.7%; Score 1461.4; DB 9; Length 1689;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1462; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 3363 AGCTGAATCTAGTGGGCCCCCTTTCAACGAGCAGACTTCAACCTCCCGCTGGGG 3422  
Db 227 ATCTGAATCTAGTGGGCCCCCTTTCAACGAGCAGACTTCAACCTCCCGCTGGGG 286  
Qy 3423 CAGAGACTCTGCCAGGCGCACTGTTGACTTGGAGTGTGGGAGGCTCAATGACAAAC 3482  
Db 287 CAGAGACTCTGCCAGGCGCACTGTTGACTTGGAGTGTGGGAGGCTCAATGACAAAC 346  
Qy 3483 TCGCGTGAACAGAACTACAGGCGCTACAGCCACTTCTGTCTTACTTGGCGGCTCA 3542  
Db 347 TCGCGTGAACAGAACTACAGGCGCTACAGCCACTTCTGTGTACTTGGCGGCTCA 406  
Qy 3543 ACCGTGAGCTGCACTGCTGAGCTCGCGCAGCTGGCGCCACTTCTGCACCAAGCTCC 3602  
Db 407 ACCGTGAGCTGCACTGCTGAGCTCGCGCAGCTGGCGCCACTTCTGCACCAAGCTCC 466  
Qy 3603 AGGGCTGTGGGCGCATTGGCGGCGTCAATGCGAGCTTGGGCTACCACTGCCCCAGC 3662  
Db 467 AGGGCTGTGGGCGCATTGGCGGCGTCAATGCGAGCTTGGGCTACCACTGCCCCAGC 526  
Qy 3663 CGTGTGCGAGTGAACCCACTTGGACTCTCGCCCTGCCACAGTACTTCTCCAGA 3722  
Db 527 CGTGTGCGAGTGAACCCACTTGGACTCTCGCCCTGCCACAGTACTTCTCCAGA 586  
Qy 3723 AGATGACGACTTCTGTGCTCTGAAGGAGCTGCAGAGCTTGGCTGTGGCGTCCGCAAGG 3782  
Db 587 AGATGACGACTTCTGTGCTCTGAAGGAGCTGCAGAGCTTGGCTGTGGCGTCCGCAAGG 646  
Qy 3783 ACTTCAACCGGCTCAAGAAAGATGAGCGCTCAGAGCTGCAGTCACTGACCTGG 3842  
Db 647 ACTTCAACCGGCTCAAGAAAGATGAGCGCTCAGAGCTGCAGTCACTGACCTGG 706  
Qy 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3902  
Db 707 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 766  
Qy 3903 CACTTTGTGAGAGCAGCCCTGTATGCAACACTGTTGAGCAGGAGACAGAACTGTG 3962



Db	767	CAC	TTTGTGAGACCGCCCTGTATGCCAACACCTGTTGAGCCAGGAGACAGAAAGCTGTG	856
Qy	3963	AGC	CTCTGGCCCTTTCTCTGACCGCGTGGCGGTGATGCGATCAGCCCTGTCTCTCCC	4022
Db	827	AGC	CTCTGGCCCTTTCTCTGACCGCGTGGCGGTGATGCGATCAGCCCTGTCTCTCCC	896
Qy	4023	CAC	CTCCCAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTCTCTGTT	4082
Db	887	CAC	CTCCCAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTCTCTGTT	946
Qy	4083	TCT	CAGGAAGTCATGCTCGAGGAGTGTGAAGTGTTCAGGTGTGTGAGAGGCGCTCA	4142
Db	947	TCT	CAGGAAGTCATGCTCGAGGAGTGTGAAGTGTTCAGGTGTGTGAGAGGCGCTCA	1006
Qy	4143	TGG	CTCTCTCTTTCTTTCGCTACCATCTGGCCAGTGCCCAACCCAGCCCTCAGGTGGCACA	4202
Db	1007	TGG	CTCTCTCTTTCTTTCGCTACCATCTGGCCAGTGCCCAACCCAGCCCTCAGGTGGCACA	1066
Qy	4203	TCT	GAGGGCAGGGTTGAGGGCCACACACACATGCTTCTGGGGTGAAGCCCTTT	4262
Db	1067	TCT	GAGGGCAGGGTTGAGGGCCACACACACATGCTTCTGGGGTGAAGCCCTTT	1126
Qy	4263	GGC	TGCCCACTCTCTCTTGGATGGGTGTGCTCCCTTATCCCAATCACTCTATACATC	4322
Db	1127	GGC	TGCCCACTCTCTCTTGGATGGGTGTGCTCCCTTATCCCAATCACTCTATACATC	1186
Qy	4323	CAAT	TCAAGAAACAACATGTGTGGCAATCTCTACACAAAGAGATGAGATTAACTGCA	4382
Db	1187	CAAT	TCAAGAAACAACATGTGTGGCAATCTCTACACAAAGAGATGAGATTAACTGCA	1246
Qy	4383	GGT	TGGGGTCTGCATTTGAGGTGCCCTATAAACCAAGAGAAAAATCTGAAAGCACAG	4442
Db	1247	GGT	TGGGGTCTGCATTTGAGGTGCCCTATAAACCAAGAGAAAAATCTGAAAGCACAG	1306
Qy	4443	GGC	AGGACACAGACACAGACCCAGGAGTCTCAAAGCACAGAGTGGCAAAACAAAC	4502
Db	1307	GGC	AGGACACAGACACAGACCCAGGAGTCTCAAAGCACAGAGTGGCAAAACAAAC	1366
Qy	4503	CCG	AGCTGAGCATCAGGACCTTGCCCTCGAATTGTCTCCAGTATTACGGTGCTCTCTC	4562
Db	1367	CCG	AGCTGAGCATCAGGACCTTGCCCTCGAATTGTCTCCAGTATTACGGTGCTCTCTC	1426
Qy	4563	TGC	CCCCCTTCCAGGGTATCTGTGGGTGCGAGCTGGGAGGCAACCATAGCCACAC	4622
Db	1427	TGC	CCCCCTTCCAGGGTATCTGTGGGTGCGAGCTGGGAGGCAACCATAGCCACAC	1486
Qy	4623	CAC	AGGATTTCTCTGAAAGTTTACAATGAGTAGCATTTTGGGGTGTAGGGTGGCGCTCC	4682
Db	1487	CAC	AGGATTTCTCTGAAAGTTTACAATGAGTAGCATTTTGGGGTGTAGGGTGGCGCTCC	1546
Qy	4683	CCA	AGGCTGCCCCCGAGCCCAACCACTCATGACTCTAAAGTGTGTTGATTAAATTTT	4742
Db	1547	CCA	AGGCTGCCCCCGAGCCCAACCACTCATGACTCTAAAGTGTGTTGATTAAATTTT	1606
Qy	4743	ATT	TATTGGAGATGTTATTATTATAGATGATATTATTGCAAGATTTCTATTCTGTATT	4802
Db	1607	ATT	TATTGGAGATGTTATTATTATAGATGATATTATTGCAAGATTTCTATTCTGTATT	1666
Qy	4803	AACA	ATAAATGCTTGCCCCAG	4825
Db	1667	AACA	ATAAATGCTTGCCCCAG	1699

RESULT	10
AX205024	
LOCUS	881 bp DNA linear PAT 30-AUG-2001
DEFINITION	Sequence 1 from Patent WO0155172.
ACCESSION	AX205024
VERSION	AX205024.1 GI:15394259
KEYWORDS	'
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS	1. (bases 1 to 881) Elson, G., Gauchat, J. F., Plun-Favreau, H., Chevalier, S. and Gascan, H.
TITLE	Isolated complex comprising a nnt-1 protein and in addition at least a clif-1 protein and/or a scntfr_g(a) protein
JOURNAL	Patent: WO 0155172-A 1 02-AUG-2001; PIERRE FABRE MEDICAMENT (FR) INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM) (FR)
FEATURES	Location/Qualifiers
source	1..881
CDS	/organism="Homo sapiens" /db_xref="taxon:9606" 174..851 /note="NNT-1" /codon_start=1 /protein_id="CAG60175.1" /db_xref="GI:15394260" /translation="MDLRAGDSWGLACLCTVLWHLPAVPAALNRRTGDPGPGSIQKTYDLTRYLEHOLARSGLVLYNLYGPPNPEPDPENPRALGAETLPRATVLEWVRSUNDKRLQTVNZEAYSHLLCYLRGLNRQAAATAEIRSLAHFCTSLQGLIGSIAGVAAALGYPLPQLPTGTEPTTPGPAHSDFELQKMDDFLWLLKELOTWLRSAKDFNRLKKKMQPPAAAVTLHLGAHG"
BASE COUNT	158 a 318 c 246 g 159 t
ORIGIN	
Query Match	10.13%; Score 525.4; DB 6; Length 881;
Best Local Similarity	99.88%; Pred. No. 2.5e-117;
Matches 526; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Qy 3363	AGCTGAACACTACCTGGGCCCCCTTTTCAACGAGCCAGACTTTCAACCTCCCGCTGGGG 3422
Db 355	ATCTGAACACTACCTGGGCCCCCTTTTCAACGAGCCAGACTTTCAACCTCCCGCTGGGG 414
Qy 3423	CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGGGAAAGCCTCAATGACAAAC 3482
Db 415	CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGGGAAAGCCTCAATGACAAAC 474
Qy 3483	TGCGGCTGACCCAGAACTACGAGGCCTACAGCCACTTCTGTGTTACTTGGCGTGCCTCA 3542
Db 475	TGCGGCTGACCCAGAACTACGAGGCCTACAGCCACTTCTGTGTTACTTGGCGTGCCTCA 534
Qy 3543	ACGTCAGGCTGCCACTGTGTAGCTGGCGCAGCTGGGCCACTTCTGCACGAGCTCC 3602
Db 535	ACGTCAGGCTGCCACTGTGTAGCTGGCGCAGCTGGGCCACTTCTGCACGAGCTCC 594
Qy 3603	AGGCGCTGCTGGGCGACATTCGGGCGTCATGCGAGCTCTGGGCTACCACTGCCCCAGC 3662
Db 595	AGGCGCTGCTGGGCGACATTCGGGCGTCATGCGAGCTCTGGGCTACCACTGCCCCAGC 654
Qy 3663	CGTGTGCGGACTGAACCCACTTGGACTCTTGGCCCTGCCACAGTGACTTCTCCAGA 3722
Db 655	CGTGTGCGGACTGAACCCACTTGGACTCTTGGCCCTGCCACAGTGACTTCTCCAGA 714
Qy 3723	AGATGACGACTTCTGGCTGCTGAAGAGCTGACACTGTGGCTGTGGGCTCGGCCAAG 3782
Db 715	AGATGACGACTTCTGGCTGCTGAAGAGCTGACACTGTGGCTGTGGGCTCGGCCAAG 774
Qy 3783	ACTTCAACCGGCTCAAGAAAGAGATGCAGCCTCCAGCAGCTCGAGTCACCTGCACTGG 3842
Db 775	ACTTCAACCGGCTCAAGAAAGAGATGCAGCCTCCAGCAGCTCGAGTCACCTGCACTGG 834
Qy 3843	GGGCTCATGGCTTCTGACTTCTGACTTCTCCTCTTGGCTCCCCCTT 3889
Db 835	GGGCTCATGGCTTCTGACTTCTGACTTCTCCTCTTGGCTCCCCCTT 881
RESULT 11	
AX205042	
LOCUS	AX205042 881 bp DNA linear PAT 30-AUG-2001
DEFINITION	Sequence 1 from Patent WO0155219.
ACCESSION	AX205042
VERSION	AX205042.1 GI:15394277
KEYWORDS	

KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 797)
AUTHORS	Chang,M.-s.
TITLE	Neurotrophic factor NNT-1
JOURNAL	Patent: US 5741772-A 1 21-APR-1998;
FEATURES	Location/Qualifiers
source	1..797
BASE COUNT	139 a 297 c 218 g 143 t
ORIGIN	/organism="unknown"
Query Match	10.3%; Score 523.4; DB 6; Length 797;
Best Local Similarity	99.8%; Pred. No. 7,7e-117;
Matches 524; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY 3363	AGCTGAACCTACCTGGCCCCCTTTCAACGAGCAGACTTCAACCCCTCCCGCTGGGG 3422
DB 271	ATCTGAACCTACCTGGCCCCCTTTCAACGAGCAGACTTCAACCCCTCCCGCTGGGG 330
QY 3423	CAGAGACTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGGAAGCCTCAATGACAAAC 3482
DB 331	CAGAGACTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGGAAGCCTCAATGACAAAC 330
QY 3483	TGGCGCTGACCCAGAACTACGAGGCCCTACAGCCACCTTCTGTGTTTACTTGGCTGGCCTCA 3542
DB 391	TGGCGCTGACCCAGAACTACGAGGCCCTACAGCCACCTTCTGTGTTTACTTGGCTGGCCTCA 450
QY 3543	ACCGTCAGGTGTCACCTGCTGAGCTGGCCGCGAGCTGGCCCACTTCTGCACCAAGCTCC 3602
DB 451	ACCGTCAGGTGTCACCTGCTGAGCTGGCCGCGAGCTGGCCCACTTCTGCACCAAGCTCC 510
QY 3603	AGGCGCTGCTGGGCGAGCATTTGCGGGCGTCAATGGCAGCTCTGGGCTACCCACTGCCCCAGC 3662
DB 511	AGGCGCTGCTGGGCGAGCATTTGCGGGCGTCAATGGCAGCTCTGGGCTACCCACTGCCCCAGC 570
QY 3663	CGCTGCCTGGGACTGAACCCACTTTGGAATCTCTGGCCCTGCCCAAGTACTTCTCCAGA 3722
DB 571	CGCTGCCTGGGACTGAACCCACTTTGGAATCTCTGGCCCTGCCCAAGTACTTCTCCAGA 630
QY 3723	AGATGGACGACTTCTGCTGCTGGAAGAGCTGCAGACCTGGCTGTGGCGCTCGGCCCAAGG 3782
DB 631	AGATGGACGACTTCTGCTGCTGGAAGAGCTGCAGACCTGGCTGTGGCGCTCGGCCCAAGG 690
QY 3783	ACTTCAACCGGCTCAAGAAAGATGACGCTTCAGAGCTTCAGAGTGCAGTCAACCTGCACCTGG 3842
DB 691	ACTTCAACCGGCTCAAGAAAGATGACGCTTCAGAGTGCAGTCAACCTGCACCTGG 750
QY 3843	GGGCTCATGGCTTCTGACTTCTGACCTTCTCCTCTTGGCTCCCC 3887
DB 751	GGGCTCATGGCTTCTGACTTCTGACCTTCTCCTCTTGGCTCCCC 795
RESULT 13	
AX392086	
LOCUS	AX392086 797 bp DNA linear PAT 23-MAR-2002
DEFINITION	Sequence 1 from Patent WO0215977.
ACCESSION	AX392086
VERSION	AX392086.1 GI:19700574
KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1
AUTHORS	Senaldi,G.
TITLE	Methods and compositions for treating ige-related disease using nn t-1 inhibitors
JOURNAL	Patent: WO 0215977-A 1 28-FEB-2002;
FEATURES	Amgen Inc. (US) Location/Qualifiers

```

source      1. .797
/organism="Homo sapiens"
/db_xref="taxon:9606"
90. .767
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD28995.1"
/db_xref="GI:19700575"
/translation="MDLRAGDSWGLACLCTVLWHLPAVPAALNRTGDPGPGSIQKTY
DLTRYLEHQLSLAGTYLNLGPPNEDFPNPRGLAETLPRATVDLEVWRSNDKLR
LTQNYEAYSHLLCYLRGLNRQATAEIRSLAHFCTSLQGLIGSIAGVMAALGYPLPQ
PLPGTEPTWTPGAHSDFLQKMDDFLLKELQTLWLSAKDFNRLKKMQPPAAAVTL
HLGAHGF"
mat_peptide 171..764
/product="Met at -27"
BASE COUNT 139 a 297 c 218 g 143 t
ORIGIN
Query Match      10.3%; Score 523.4; DB 6; Length 797;
Best Local Similarity 99.8%; Pred. No. 7.7e-117;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422
Db      |||||
QY 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 330
Db      |||||
QY 3423 CAGAGACTCTGCCAGGGCCACTGTGACTTGGAGTGTGGGAAAGCCCTCAATGACAAAC 3482
Db      |||||
QY 331 CAGAGACTCTGCCAGGGCCACTGTGACTTGGAGTGTGGGAAAGCCCTCAATGACAAAC 390
Db      |||||
QY 3483 TCGGGCTGACCCAGAACTACGAGGCTTACAGCCACCTCTGTGTTACTTGGCTGCTCA 3542
Db      |||||
QY 391 TCGGGCTGACCCAGAACTACGAGGCTTACAGCCACCTCTGTGTTACTTGGCTGCTCA 450
Db      |||||
QY 3543 ACCGTGAGCTGCCACTGTGAGCTGCGCGCAGCTTCTGCACAGCTCC 3602
Db      |||||
QY 451 ACCGTGAGCTGCCACTGTGAGCTGCGCGCAGCTTCTGCACAGCTCC 510
Db      |||||
QY 3603 AGGGCTGCTGGGAGCAATGCGGGCGTATGGCAGCTCTGGGCTACCCACTGCCCGCAGC 3662
Db      |||||
QY 511 AGGGCTGCTGGGAGCAATGCGGGCGTATGGCAGCTCTGGGCTACCCACTGCCCGCAGC 570
Db      |||||
QY 3663 CGTGCCTGGGACTGAACCCACTTGGACTCTGCGCCCTGCCACAGTGAATCTCTCCAGA 3722
Db      |||||
QY 571 CGTGCCTGGGACTGAACCCACTTGGACTCTGCGCCCTGCCACAGTGAATCTCTCCAGA 630
Db      |||||
QY 3723 AGATGAGCACTTCTGGCTGTGAAGAGCTGCAGACTGCTGGGCTCGGCCCAAGG 3782
Db      |||||
QY 631 AGATGAGCACTTCTGGCTGTGAAGAGCTGCAGACTGCTGGGCTCGGCCCAAGG 690
Db      |||||
QY 3783 ACTTCAACCGGCTCAAGAGAGATGACGCTTCCAGCAGCTGCACTGCACTCC 3842
Db      |||||
QY 691 ACTTCAACCGGCTCAAGAGAGATGACGCTTCCAGCAGCTGCACTGCACTCC 750
Db      |||||
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGCTCTTCTGCTCTCCCC 3887
Db      |||||
QY 751 GGGCTCATGGCTTCTGACTTCTGACTTCTGCTCTTCTGCTCTCCCC 795
Db      |||||

RESULT 14
AF176911
LOCUS      AF176911
DEFINITION Homo sapiens neurotrophin-1/B-cell stimulating factor-3 mRNA,
complete cds.
ACCESSION AF176911
VERSION    AF176911.1
KEYWORDS   AF176911.1 GI:6007640
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 797)
AUTHORS    Senaldi,G., Varnum,B.C., Sarmiento,U., Lile,J., Starnes,C.,

```

```

TITLE      Scully,S., Guo,J., Elliott,G., McNinch,J., Shalee,C.L.,
Freeman,D., Manu,F., Simonet,W.S., Boone,T. and Chang,M.S.
Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the
11-6 family
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)
MEDLINE    99432254
PUBMED     10500198
REFERENCE  2 (bases 1 to 797)
AUTHORS    Senaldi,G., Varnum,B., Sarmiento,U., Lile,J., Starnes,C.,
Scully,S., Guo,J., Elliott,G., McNinch,J., Freeman,D., Shalee,C.,
Manu,F., Simonet,W.S., Boone,T. and Chang,M.-S.
DIRECT SUBMISSION
JOURNAL    Submitted (11-AUG-1999) Amgen, Inc., One Amgen Center Drive,
Thousand Oaks, CA 91320, USA
FEATURES   Location/Qualifiers
            1..797
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            90..767
            /codon_start=1
            /product="neurotrophin-1/B-cell stimulating factor-3"
            /protein_id="AAF00991.1"
            /db_xref="GI:6007641"
            /translation="MDLRAGDSWGLACLCTVLWHLPAVPAALNRTGDPGPGSIQKTY
DLTRYLEHQLSLAGTYLNLGPPNEDFPNPRGLAETLPRATVDLEVWRSNDKLR
LTQNYEAYSHLLCYLRGLNRQATAEIRSLAHFCTSLQGLIGSIAGVMAALGYPLPQ
PLPGTEPTWTPGAHSDFLQKMDDFLLKELQTLWLSAKDFNRLKKMQPPAAAVTL
HLGAHGF"
BASE COUNT 139 a 297 c 218 g 143 t
ORIGIN

```

```

Query Match      10.3%; Score 523.4; DB 9; Length 797;
Best Local Similarity 99.8%; Pred. No. 7.7e-117;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422
Db      |||||
QY 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 330
Db      |||||
QY 3423 CAGAGACTCTGCCAGGGCCACTGTGACTTGGAGTGTGGGAAAGCCCTCAATGACAAAC 3482
Db      |||||
QY 331 CAGAGACTCTGCCAGGGCCACTGTGACTTGGAGTGTGGGAAAGCCCTCAATGACAAAC 390
Db      |||||
QY 3483 TCGGGCTGACCCAGAACTACGAGGCTTACAGCCACCTCTGTGTTACTTGGCTGCTCA 3542
Db      |||||
QY 391 TCGGGCTGACCCAGAACTACGAGGCTTACAGCCACCTCTGTGTTACTTGGCTGCTCA 450
Db      |||||
QY 3543 ACCGTGAGCTGCCACTGTGAGCTGCGCGCAGCTTCTGCACAGCTCC 3602
Db      |||||
QY 451 ACCGTGAGCTGCCACTGTGAGCTGCGCGCAGCTTCTGCACAGCTCC 510
Db      |||||
QY 3603 AGGGCTGCTGGGAGCAATGCGGGCGTATGGCAGCTCTGGGCTACCCACTGCCCGCAGC 3662
Db      |||||
QY 511 AGGGCTGCTGGGAGCAATGCGGGCGTATGGCAGCTCTGGGCTACCCACTGCCCGCAGC 570
Db      |||||
QY 3663 CGTGCCTGGGACTGAACCCACTTGGACTCTGCGCCCTGCCACAGTGAATCTCTCCAGA 3722
Db      |||||
QY 571 CGTGCCTGGGACTGAACCCACTTGGACTCTGCGCCCTGCCACAGTGAATCTCTCCAGA 630
Db      |||||
QY 3723 AGATGAGCACTTCTGGCTGTGAAGAGCTGCAGACTGCTGGGCTCGGCCCAAGG 3782
Db      |||||
QY 631 AGATGAGCACTTCTGGCTGTGAAGAGCTGCAGACTGCTGGGCTCGGCCCAAGG 690
Db      |||||
QY 3783 ACTTCAACCGGCTCAAGAGAGATGACGCTTCCAGCAGCTGCACTGCACTCC 3842
Db      |||||
QY 691 ACTTCAACCGGCTCAAGAGAGATGACGCTTCCAGCAGCTGCACTGCACTCC 750
Db      |||||
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGCTCTTCTGCTCTCCCC 3887
Db      |||||
QY 751 GGGCTCATGGCTTCTGACTTCTGACTTCTGCTCTTCTGCTCTCCCC 795
Db      |||||

RESULT 15

```

Sun Feb 2 08:32:00 2003

us-09-931-704-3.rge

AY049779  
 LOCUS AY049779 680 bp mRNA linear PRI 23-OCT-2001  
 DEFINITION Homo sapiens cardiotrophin-like cytokine (CLC) mRNA, complete cds.  
 ACCESSION AY049779  
 VERSION AY049779.1 GI:16356642  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 680)  
 Hu.X., Xu.Y., Zhang, B., Peng, X., Yuan, J. and Qiang, B.  
 Direct Submission  
 TITLE Submitted (30-JUL-2001) Department of Biochemistry, Institute of  
 JOURNAL Basic Medical Science, Chinese Academy of Medical Sciences, 5 Dong  
 Dan San Tiao, Beijing 100005, P.R. China

FEATURES  
 source  
 1..680 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..680 /gene="CLC"  
 3..680 /gene="CLC"  
 /note="neurotrophin-1/B-cell stimulating factor-3"  
 /codon\_start=1  
 /product="cardiotrophin-like cytokine"  
 /protein\_id="AAL15436.1"  
 /db\_xref="GI:16356643"  
 /translations="MDLRAGDSWGLACLCTVLWHLPAVPAALNRTGDPGSPISQKTY  
 DLTRYEHQRLSAGTYLVNLYGPPNEPDPNPRLEAETLPRAVDLEWVRSNDKLR  
 LTONYRAYSHLLCYLRGLNQAAATAEIARRSLAHFCTSLQGLGSIAGVMAALGYPLPQ  
 PLFGTEPTWTPGPAHSDFLQMDDFLWLLKELQTLWLSAKDFNRLKKMQPPAAAVIL  
 HLGAGGF"  
 BASE COUNT 125 a 239 c 191 g 125 t  
 ORIGIN

Query Match 9.7%; Score 495.4; DB 9; Length 680;  
 Best Local Similarity 99.8%; Pred. No. 5.4e-110;  
 Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3363 AGCTGAACCTACTGGGCCCCCTTTCAACAGAGCCAGACTTCAACCCCTCCCGCCCTGGGGG 3422  
 DB 184 ATCTGAACCTACTGGGCCCCCTTTCAACAGAGCCAGACTTCAACCCCTCCCGCCCTGGGGG 243  
 QY 3423 CAGAGACTCTGCCAGGCGCACTGTTGACTTGGAGGTGTGGAGGCTCAATGACAAAC 3482  
 DB 244 CAGAGACTCTGCCAGGCGCACTGTTGACTTGGAGGTGTGGAGGCTCAATGACAAAC 303  
 QY 3483 TGGGCTGACCCAGAACTACGAGGCTACAGCCACCTTCTGTGTTACTTGGTGGCTCA 3542  
 DB 304 TGGGCTGACCCAGAACTACGAGGCTACAGCCACCTTCTGTGTTACTTGGTGGCTCA 363  
 QY 3543 ACCGTGAGGCTGCCACTGTGAGCTGCGCGCAGAGCTGGGCCACTTCTGCACAGGCTCC 3602  
 DB 364 ACCGTGAGGCTGCCACTGTGAGCTGCGCGCAGAGCTGGGCCACTTCTGCACAGGCTCC 423  
 QY 3603 AGGGCTGTGGGAGCAGCATTTGGGGGGTGTGGAGCTTGGGCTACCCACTGCCCCAGC 3662  
 DB 424 AGGGCTGTGGGAGCAGCATTTGGGGGGTGTGGAGCTTGGGCTACCCACTGCCCCAGC 483  
 QY 3663 CGCTGCTGGGACTGAACCCACTTGGAGCTCTGGCCCTGCCACAGTGAATTCCTCCAGA 3722  
 DB 484 CGCTGCTGGGACTGAACCCACTTGGAGCTCTGGCCCTGCCACAGTGAATTCCTCCAGA 543  
 QY 3723 AGATGACGACTTCTGGTGTGAAGAGCTGCAGAGCTGGCTGTGGCGTCCGCCAAGG 3782  
 DB 544 AGATGACGACTTCTGGTGTGAAGAGCTGCAGAGCTGGCTGTGGCGTCCGCCAAGG 603  
 QY 3783 ACTTCAACGGGCTCAAGAAAGATGAGCTCCAGAGCTGCAGTCAACCTGACCTGG 3842  
 DB 604 ACTTCAACGGGCTCAAGAAAGATGAGCTCCAGAGCTGCAGTCAACCTGACCTGG 663

QY 3843 GGGCTCATGGCTTCTGA 3859  
 DB 664 GGGCTCATGGCTTCTGA 680  
 Search completed: February 1, 2003, 18:01:12  
 Job time : 14154.7 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 10332.6 Seconds  
(without alignments)  
7973.437 Million cell updates/sec

Title: US-09-931-704-3  
Perfect score: 5087  
Sequence: 1 aacctgcagtgaggctggc.....cctttgtaagtctctcca 5087

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

EST:\*

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estmu.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estio.\*
- 8: em\_hic.\*
- 9: gb\_est1.\*
- 10: gb\_est2.\*
- 11: gb\_hic.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estom.\*
- 17: gb\_gss.\*
- 18: em\_gss\_hum.\*
- 19: em\_gss\_inv.\*
- 20: em\_gss\_pln.\*
- 21: em\_gss\_vrt.\*
- 22: em\_gss\_fun.\*
- 23: em\_gss\_mam.\*
- 24: em\_gss\_mus.\*
- 25: em\_gss\_other.\*
- 26: em\_gss\_pro.\*
- 27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	860.2	16.9	955	9	AL543945
C 2	859.8	16.9	888	9	AL570325
C 3	656.6	12.9	681	10	AW978561
4	612	12.0	691	12	BF213570
5	567.4	11.2	915	12	BF035982
6	512.4	10.1	514	12	BF058969

7	488.6	9.6	1157	14	BQ940483
C 8	487	9.6	487	10	BE463575
C 9	479.4	9.4	493	9	AI040033
10	471.4	9.3	953	12	BQ437538
11	470.4	9.2	887	14	BQ948158
C 12	465.4	9.1	467	12	BF002167
C 13	465.4	8.9	454	12	BF513712
C 14	450.4	8.9	457	10	AW044541
C 15	446.2	8.8	462	9	AI916548
C 16	429.4	8.6	465	9	AA430602
C 17	429.4	8.4	432	9	AA430517
C 18	422.4	8.3	432	9	AA278936
C 19	422.2	8.3	433	9	AI765382
C 20	407.4	8.0	409	9	AA283037
C 21	405	8.0	411	9	AI752562
C 22	399	7.8	407	9	AI752561
C 23	383.8	7.5	1053	12	BG164929
C 24	375	7.4	381	12	BF056328
C 25	374.2	7.4	452	9	AA283056
C 26	373.8	7.3	401	12	BF881796
C 27	372.6	7.3	396	12	BF849577
C 28	355.8	7.0	360	9	AA284882
C 29	352.4	6.9	354	9	AA830376
C 30	351	6.9	471	12	BF852246
C 31	350.4	6.9	594	14	BM763333
C 32	346.8	6.8	512	9	AI390475
C 33	343.2	6.7	356	9	AI769263
C 34	342.2	6.7	476	10	BE632644
C 35	342.2	6.6	622	12	BF852748
C 36	337.4	6.6	580	14	BM848189
C 37	333.2	6.6	448	12	BF853441
C 38	331.6	6.5	497	12	BF378449
C 39	331.4	6.5	573	14	BM840863
C 40	320	6.3	321	9	AI474292
C 41	319.4	6.3	346	9	AI868273
C 42	319.2	6.3	458	13	BM363136
C 43	316.4	6.2	351	9	AI571829
C 44	311.6	6.1	468	12	BF369558
C 45	311.2	6.1	457	12	BF750407

#### ALIGNMENTS

RESULT 1  
AL543945  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

AL543945  
AL543945 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0DI004YM15 5  
prime, mRNA sequence.  
AL543945  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
1 (bases 1 to 955)  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers  
1..955  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DI004YM15"  
/clone\_lib="LTI\_NFL006\_PL2"  
/tissue\_type="placenta"  
/notes="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com

BASE COUNT 168 a 323 c 286 g 198 t  
ORIGIN  
Query Match 16.9%; Score 860.2; DB 9; Length 955;  
Best Local Similarity 99.2%; Pred. No. 6.4e-186;  
Matches 896; Conservative 0; Mismatches 3; Indels 4; Gaps 3;  
QY 3363 AGCTGAACCTACCTGGCCGCCCTTTCAACGAGCGACACTTCAACCCCTCCCGCTGGGG 3422  
DB 53 ATCTGAACCTACCTGGCCGCCCTTTCAACGAGCGACACTTCAACCCCTCCCGCTGGGG 112  
QY 3423 CAGAGACTTGGCCAGGCGCACTGTTGACTTGGAGGTGCGGAAGCTCAATGACAAAC 3482  
DB 113 CAGAGACTTGGCCAGGCGCACTGTTGACTTGGAGGTGCGGAAGCTCAATGACAAAC 170  
QY 3483 TGGCGCTGACCCAGAACTACGAGCGCTTACAGCCACTTCTGTGTTACTTGGTGGCTCA 3542  
DB 171 TGGCGCTGACCCAGAACTACGAGCGCTTACAGCCACTTCTGTGTTACTTGGTGGCTCA 230  
QY 3543 ACCGTGAGGTGCACTGCTGAGTGGCGCGAGCTGGCCACTTCTGCACGAGCTCC 3602  
DB 231 ACCGTGAGGTGCACTGCTGAGTGGCGCGAGCTGGCCACTTCTGCACGAGCTCC 290  
QY 3603 AGGGCTGCTGGCGAGCACTTGGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCCAGC 3662  
DB 291 AGGGCTGCTGGCGAGCACTTGGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCCAGC 350  
QY 3663 CGCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCACAGTACTTCTCTCCAGA 3722  
DB 351 CGCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCACAGTACTTCTCTCCAGA 410  
QY 3723 AGATGAGCACTTCTGCTGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAG 3782  
DB 411 AGATGAGCACTTCTGCTGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAG 470  
QY 3783 ACTTCAACCGGCTCAAGAAAGATGAGAGCTCCAGAGCTGCACTGCACTGCACTGCTG 3842  
DB 471 ACTTCAACCGGCTCAAGAAAGATGAGAGCTCCAGAGCTGCACTGCACTGCACTGCTG 530  
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCT 3902  
DB 531 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCT 590  
QY 3903 CACTTTGTGAGAGCGAGCCCTGTATGCCAACACCTGTTGAGCGAGGAGACAGAACTGTG 3962  
DB 591 CACTTTGTGAGAGCGAGCCCTGTATGCCAACACCTGTTGAGCGAGGAGACAGAACTGTG 650  
QY 3963 AGCCTCTGGCCCTTTCTCTGGACCGGCTGGGGCTGTGATCGATCAGCCCTGTCTCTCTCC 4022  
DB 651 AGCCTCTGGCCCTTTCTCTGGACCGGCTGGGGCTGTGATCGATCAGCCCTGTCTCTCTCC 710  
QY 4023 CACTTCCAAAGGCTTACCGAGCT--GGGAGGAGGTACAGTAGGCGCTGCTGCTGCTGCTG 4081  
DB 711 CACTTCCAAAGGCTTACCGAGCTTGGGAGGAGGTACAGTAGGCGCTGCTGCTGCTGCTGCTG 770  
QY 4082 TTCTACAGGAAGTATGCTCGAGGAGGTGTGAAGTGGTTCAGGTGGTGGTGGTGGTGGTGGTGGT 4141  
DB 771 TTCTACAGGAAGTATGCTCGAGGAGGTGTGAAGTGGTTCAGGTGGTGGTGGTGGTGGTGGT 830  
QY 4142 ATGGCTCCTGCTTCTGCTTACCACTTGGCCAGTGGCCACCGAGCCCTCAGGTGGGAC 4201  
DB 831 ATGGCTCCTGCTTCTGCTTACCACTTGGCCAGTGGCCACCGAGCCCTCAGGTGGGAC 890  
QY 4202 ATCTGGAGGCA--GGGTTGAGGGGCCACACACATGCTCTTCTGGGTTGAAGCCCT 4260

DB 891 ATCTGGAGGCGAGCGGCTTGTAGGGGCCACACACACATGCTTCTGGGTTGAAGCCCTT 950  
QY 4261 TTG 4263  
DB 951 TGG 953  
RESULT 2  
AL570325/c 888 bp mRNA linear EST 16-FEB-2001  
LOCUS AL570325 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0DI004YM15 3  
DEFINITION prime, mRNA sequence.  
ACCESSION AL570325  
VERSION AL570325.1 GI:12926520  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 888)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
Location/Qualifiers  
1..888  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DI004YM15"  
/tissue\_type="placenta"  
/note="Vector: pCMVSPORT 6; Site: 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"  
BASE COUNT 204 a 233 c 247 g 199 t  
ORIGIN  
Query Match 16.9%; Score 859.8; DB 9; Length 888;  
Best Local Similarity 99.0%; Pred. No. 7.8e-186;  
Matches 880; Conservative 5; Mismatches 2; Indels 2; Gaps 2;  
QY 3884 CCCCTTCAACCCCTGCTCCCACTTTGTGAGAGCGAGCCCTGTATG-CCAACACCTGTGA 3942  
DB 888 CCCYTTCAACCCCTGCTCCCACTTTGTGAGAGCGAGCCCTGTATGCCCTACACCTGTGA 829  
QY 3943 GCCAGAGACAGAACTGTGAGCTCTGGCCCTTCTGACCGGCTGGGCGTGTGATGC 4002  
DB 828 GCCAGAGACAGAACTGTGAGCTTTTGCCCTTTCTGACCGGCTGGGCGTGTGATGC 769  
QY 4003 GATCAGCCCTGTCTCTCCCACTCCCAAGGCTTACCAGCTGGGGAGGAGTACAGT 4062  
DB 768 GATCAGCCCTGTCTCTCCCACTCCCAAGGCTTACCAGCTGGGGAGGAGTACAGT 709  
QY 4063 AGGCCCTGTCTCTCTCTCTCTACAGGAAGTATGCTCAGGAGGTGTGAAGTGGTTCA 4122  
DB 708 AGGCCCTGTCTCTCTCTCTCTACAGGAAGTATGCTCAGGAGGTGTGAAGTGGTTCA 649  
QY 4123 GGTGTGTGACAGCGCTCATGCT 4182  
DB 648 GGTGTGTGACAGCGCTCATGCT 589  
QY 4183 CCAGCCCTCAGTGTGACATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4242

Db 588 .CAGCCCTCAGGTGGACATCTCGAGGGCAGGGGTGAGGGGCCACACACATGCCC 529  
QY 4243 TTTCTGGGTGAAGCCCTTTGGCTGCCCACTCTCTTTGGATGGGTGTGCTCCCTTATC 4302  
Db 528 TTTCTGGGTGAAGCCCTTTGGCTGCCCACTCTCTTTGGATGGGTGTGCTCCCTTATC 469  
QY 4303 CCAGAACTACTATACATCAATTCAGGAAACAAACATGTTGCAATTTACACAAAA 4362  
Db 468 CCCAAATCACTCTATACATCAATTCAGGAAACAAACATGTTGCAATTTACAC-AAAA 410  
QY 4363 GAGATGAGATTAAACAGTGCAGGGTTGGGTCTGCAATTTGGAGTGCCTTATAAACACAGAAG 4422  
Db 409 GAGATGAGATTAAACAGTGCAGGGTTGGGTCTGCAATTTGGAGTGCCTTATAAACACAGAAG 350  
QY 4423 AGAAATACTGAAACAGCAGGGGAGGAGACAGACAGACAGACAGACAGACAGACAGACAG 4482  
Db 349 AGAAATACTGAAACAGCAGGGGAGGAGACAGACAGACAGACAGACAGACAGACAGACAG 290  
QY 4483 CACAGAGTGGCAACAAACCCAGAGCTGAGCATCAGGACCTTGCTCGAATTTGCTTCCA 4542  
Db 289 CACAGAGTGGCAACAAACCCAGAGCTGAGCATCAGGACCTTGCTCGAATTTGCTTCCA 230  
QY 4543 GTATTACGGTGCCTTCTCTGCCCCCTTTCCAGGGTATCTGTTGGTTCAGGCTGGG 4602  
Db 229 GTATTACGGTGCCTTCTCTGCCCCCTTTCCAGGGTATCTGTTGGTTCAGGCTGGG 170  
QY 4603 GAGGGCAACCATAGCCACACACAGGATTTCTGAAAGTTTACAAATGCAAGTATGCTTTG 4662  
Db 169 GAGGGCAACCATAGCCACACACAGGATTTCTGAAAGTTTACAAATGCAAGTATGCTTTG 110  
QY 4663 GGGTGTAGGTGGAGCTCCCAAGGCCCTGCCCCCAGGCCACCCACCATCATGACTCTA 4722  
Db 109 GGGTGTAGGTGGAGCTCCCAAGGCCCTGCCCCCAGGCCACCCACCATCATGACTCTA 50  
QY 4723 AGTGTGTGATTAAATATTATTATTATTGGAGATGTTATTATTAGATG 4771  
Db 49 AGMTGTGTGATTAAATATTATTATTATTGGAGATGTTATTATTAGATG 1

RESULT 3  
AW978561/c 681 bp mRNA linear EST 02-JUN-2000  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AW978561  
EST390670 MAGE resequences, MAGP Homo sapiens cDNA, mRNA sequence.  
AW978561.1 GI:8169828  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt  
I-E, Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and  
Quackenbush, J.  
Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray  
Unpublished (2000)  
Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 394  
Seq primer: Forward.  
Location/Qualifiers  
1. .681  
/organism="Homo sapiens"  
/db xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGP"  
/note="Vector: pBluescriptSKm" 153 a 94 c 248 g 185 t 1 others

FEATURES  
source  
BASE COUNT  
ORIGIN

Query Match 12.9%; Score 656.6; DB 10; Length 681;  
Best Local Similarity 99.3%; Pred. No. 2.3e-139;  
Matches 670; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 190 TGATACCTAAACCGACCAAGTCAAGCCCTCCCAACTCACCTCTGCGCTGCCAGACCTCA 249  
Db 681 TAAATACCTAAACCGACCAATTTACAGGCTTTCCAACTCACCTCTGCGCTGCCAGACCTCA 622  
QY 250 CCACATCCTTTG-TGGAGTCAAACTCAACCGCCTAAATCAACCAATCCCAAGTCTAAA 308  
Db 621 CCACATCCTTTGNTGGAGTCAAACTCAACCGCCTAAATCAACCAATCCCAAGTCTAAA 562  
QY 309 CTAATCTGAAACTTTTAAAGTAACCCAGTCTCTTAAACCTAACCTAGGCCCAATGCAATTA 368  
Db 561 CTAATCTGAAACTTTTAAAGTAACCCAGTCTCTTAAACCTAACCTAGGCCCAATGCAATTA 502  
QY 369 TATCTACCTTAGCCAAACCTTAAGTCTGCTTTGCCAGTCCAAAGTGTCCAGTGAATCTCTCA 428  
Db 501 TATCTACCTTAGCCAAACCTTAAGTCTGCTTTGCCAGTCCAAAGTGTCCAGTGAATCTCTCA 442  
QY 429 CTTTGGTCTCACTGAAATCCCAAGAAAGCATATTTCCCACTGCCACATCCCTCTCT 488  
Db 441 CTTTGGTCTCACTTAAATCCCAAGAAAGCATATTTCCCACTGCCACATCCCTCTCT 382  
QY 489 ACAGCACCACCAACCTGCGCTCTGGACTCTCTGGTATCTCTGGGATGTCCTCAAACTCTGCAATG 548  
Db 381 ACAGCACCACCAACCTGCGCTCTGGACTCTCTGGTATCTCTGGGATGTCCTCAAACTCTGCAATG 322  
QY 549 CCATCAGCACAAGCCCGAGTCTGTAATGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 608  
Db 321 CCATCAGCACAAGCCCGAGTCTGTAATGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 262  
QY 609 GCAGGCTGTAGGAAAGCCCTCAATTTGAAAGTCCAACTTTTCCCACTGCCACATCCCAAGACGG 668  
Db 261 GCAGGCTGTAGGAAAGCCCTCAATTTGAAAGTCCAACTTTTCCCACTGCCACATCCCAAGACGG 202  
QY 669 GGTGAACCTCCACACTGCCACCTGCTTCCCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 728  
Db 201 GGTGAACCTCCACACTGCCACCTGCTTCCCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 142  
QY 729 CCCACCTCTACACTTCCACACTCAGGAATCAATCTCTAGAAATATACCAAACTAGACCC 788  
Db 141 CCCACCTCTACACTTCCACACTCAGGAATCAATCTCTAGAAATATACCAAACTAGACCC 82  
QY 789 CCATAAGGAGCCGACCCCTAGTGTGCTTAACCTTATACCTTCTCTCTCTCTCTCTCTCTCTCTCT 848  
Db 81 CCATAAGGAGCCGACCCCTAGTGTGCTTAACCTTATACCTTCTCTCTCTCTCTCTCTCTCTCTCT 22  
QY 849 GTTCTTGGCGGCGCG 863  
Db 21 GTTCTTGGCGGCGCG 7

RESULT 4  
BF213570 691 bp mRNA linear EST 06-NOV-2000  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BF213570  
G01845370F1 NIH\_MGC\_55 Homo sapiens cDNA clone IMAGE:4070560 5',  
mRNA sequence.  
BF213570  
BF213570.1 GI:11107156  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 691)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC

Qy	4327	TCAGGAACAACATGGTGGCAATTCTACACAAAAGAGATGAGATTAA	CAGTGCAGG	4386
Db	598	TCAGGAACCAACATGTTGGAATTTACACAAAAGAGATGAGATTAA	CAGTGCAGG	657
Qy	4387	TGGGGTCTGCATTGGAGGTGCCCTATAAC	CAGA 4420	
Db	658	TGGGGTCTGATTGGAGGGGCTATAAC	AGA 691	

3  
2  
3  
  
2  
1

RESULT_5	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
LOCUS	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
DEFINITION	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
ACCESSION	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
VERSION	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
KEYWORDS	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
SOURCE	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
ORGANISM	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
REFERENCE	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
AUTHORS	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
TITLE	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
JOURNAL	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000
COMMENT	BF035982	BF035982	915 bp	mRNA	linear	EST 20-OCT-2000

0000

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
CDNA Library Arrayed by: Incyte Genomics, Inc.  
DNA Sequencing by: MGC clone distribution information can be  
Clone Distribution: MGC clone distribution information can be

<http://image.llnl.gov>

High quality sequence stop: 615.

FEATURES  
source  
1. .915  
LOCATION/QUANTITIES

```
/db xref="taxon:9606"
```

```
/clone lib="NIH MGC_66"
```

```
/lab host="DH10B (phage-resistant)"
```

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.0 kb.  
" **Technologies.** Library constructed by

BASE COUNT	184 a	283 c	230 g	130 c
ORIGIN				

Query Match 11.2%: Score 567.4: DB 12: Length 915:

Best Local Similarity	87.7%	Pred: NO. 6.2E-119;
Matches 701: Conservative	0	Mismatches 86:
		Indels 12: Gaps 7:

3638 GCTCTGGGGTACCCCACTGCCCCAGCCGGTGGCTGGGA CTGAACCCCA CTCTGGACTCTCTGGC 3697

64

2500 CATTCCCACTCACTTCCTCCAGACGATCGACGACTTCTGGCTGCTGAAGAGCTGCAG 3757

124

Q. 276 ACCCCTCCACCAACCCTTCAAACCCCTCACCAAGCATGCGCTCCA 3817

183

3977

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040



```

Db 244 TCCTCTCCCTTCAAACCTCTCCACCTTTGTGAGAGCAGCCCTGTATGCCAACACCT 303
QY 3938 GTTGAGCCAGGAGACAGAAAGCTGTGAGCTCTGCGCCCTTTCTTGAGACCGCTGGCGTCT 3997
Db 304 GTTGAGCCAGGAGACAGAAAGCTGTGAGCTCTGCGCCCTTTCTTGAGACCGCTGGCGTCT 363
QY 3998 GATGATCAGCCCTGTCTCTCCACCTCTCCACCTCTCCAAAGGTCTACCGAGCTGGGAGAGGT 4057
Db 364 GATGATCAGCCCTGTCTCTCCACCTCTCCAAAGGTCTACCGAGCTGGGAGAGGT 423
QY 4058 ACAGTAGGCCCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4117
Db 424 ACAGTAGGCCCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 483
QY 4118 GTTACAGTTGTGTGACAGGCGCTCATGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4176
Db 484 GTTACAGTTGTGTGACAGGCGCTCATGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 543
QY 4177 GCCACCCAGCCCTCAGGTGGACATCTGAGGGGAGGGGTTGAGGGGCCACCCACCA 4236
Db 544 GCCACCCCA-CCCTCAGGTGGACATCTGAGGGGAGGGGTTGAGGGGCCACCCAC- 600
QY 4237 CATGCTTCTGGGGTGAAGCCCTTTGGCTGCCCACTCTCTCTCTCTCTCTCTCTCTCTCT 4292
Db 601 ATGCTTCTGGGGTGAAGCCCTTTGGATGGCCCCCACTCTCTCTCTCTCTCTCTCTCTCTCT 660
QY 4293 TTCCCTTATCCCAATCACTCTATACATCCCAATTCAGGAACAAACATGTTGGCAATTC 4352
Db 661 TTCCCTTATCCCAATCACTCTATACATTC--ATTGAGAACCACTGTTGGCAATTCATCC 718
QY 4353 TACACAAAAGAGTAGATTAACAGTGGAGGTTGGGTTGCTGCAATGGAGTGGCCCTAT 4412
Db 719 AAAAGAGATAGATTAAACGGGAGTGGGCTCTCTCTGAGGGCCATAACCGAGAAAACCTG 778
QY 4413 AAACAGAGAGAAATAC 4431
Db 779 AACCTGGGGCAACAC 797

RESULT 6
LOCUS BF058969 514 bp mRNA linear EST 16-OCT-2000
DEFINITION 7k36e10.xl NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3477595 3',
mRNA sequence.
ACCESSION BF058969
VERSION BF058969.1 GI:10812865
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LENL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 493.
Location/Qualifiers
1. .514
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="IMAGE:3477595"
/clone_lib="NCI CGAP Ov18"
/tissue_type="fibrotheoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGGCGGACATTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 96 a 195 c 112 g 111 t
ORIGIN

Query Match 10.1%; Score 512.4; DB 12; Length 514;
Best Local Similarity 99.8%; Pred. No. 2.2e-106;
Matches 513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 856 GGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 915
Db 1 GGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 60
QY 916 ACATGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 975
Db 61 ACATGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
QY 976 AGAAGCTAGGTAACAAGTGAACACAGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1035
Db 121 AGAAGCTAGGTAACAAGTGAACACAGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 180
QY 1036 GTCTCTCCATGACACACATATACATGTTGGACACACAGTGGGACACATGTTGGGACAC 1095
Db 181 GTCTCTCCATGACACACATATACATGTTGGACACACAGTGGGACACATGTTGGGACAC 240
QY 1096 ACTCTCTCAGTGCACACAGATCCATCTCAAGTATCTACTGATAGACACTCATGCGGTG 1155
Db 241 ACTCTCTCAGTGCACACAGATCCATCTCAAGTATCTACTGATAGACACTCATGCGGTG 300
QY 1156 CCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1215
Db 301 CCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
QY 1216 TTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1275
Db 361 TTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
QY 1276 CCCAAGTGGGACAGACACCTGAGGGGCTGCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1335
Db 421 CCCAAGTGGGACAGACACCTGAGGGGCTGCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
QY 1336 GCGTCTCATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1369
Db 481 GCGTCTCATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 514

RESULT 7
LOCUS BF040483 1157 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8864294 Lupski sciatic nerve Homo sapiens cDNA clone
IMAGE:6197786 5', mRNA sequence.
ACCESSION BF040483
VERSION BF040483.1 GI:22355961
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1157)
AUTHORS NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

```



```

Db 307 GACCTTGCTCGAATTGCTCCAGATTAACAGTGGCTCTCTCTGCGCCCTTTCCAGG 248
Qy 4579 GTATCTGGGTGTCAGAGCTGGGAGGGCAACCATAGCCACACACAGGATTTCTCGAA 4638
Db 247 GTATCTGGGTGTCAGAGCTGGGAGGGCAACCATAGCCACACACAGGATTTCTCGAA 188
Qy 4639 AGTTTACATGAGTAGATTTTGGGGTGTAGGGTGGGAGCTGCCAAGGGCCCTGCCCCC 4698
Db 187 AGTTTACATGAGTAGATTTTGGGGTGTAGGGTGGGAGCTGCCAAGGGCCCTGCCCCC 128
Qy 4699 CAGCCCCACCCACTCATGACTCTAAGTGTGTGTATTATTAATTTATTTATTTGGAGATGT 4758
Db 127 CAGCCCCACCCACTCATGACTCTAAGTGTGTGTATTATTAATTTATTTATTTGGAGATGT 68
Qy 4759 TATTTATTAGATGATTTATTTATTCAGAAATTTCTATTTCTGTATTAACAAATAAAATGCTT 4818
Db 67 TATTTATTAGATGATTTATTTATTCAGAAATTTCTATTTCTGTATTAACAAATAAAATGCTT 8
Qy 4819 GCCCAG 4825
Db 7 GCCCAG 1

RESULT 9
AI040033/c
LOCUS 493 bp mRNA linear EST 24-SEP-1998
DEFINITION ox97h03.x1 Soares senescent fibroblasts_NbHSF Homo sapiens cDNA
clone IMAGE:1664309 3', mRNA sequence.
ACCESSION AI040033
VERSION AI040033.1 GI:3279227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 493)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 954 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 456.
FEATURES
Location/Qualifiers
1..493
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:1664309"
/clone_lib="Soares_senescent_fibroblasts_NbHSF"
/tissue_type="senescent_fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/notes="vector: pT7T3D (Pharmacia) with a modified
polylinker V TYPE: phagemid; Site 1: Not I; Site 2: Eco RI
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTTACCAATCTGAAGTGGGAGGGCGGCGGATTTTTTTTTTTTTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Patima Bonaldo."
BASE COUNT 121 a 113 c 112 g 147 t
ORIGIN

Query Match 9.4%; Score 479.4; DB 9; Length 493;
Best Local Similarity 99.8%; Pred. No. 7.6e-99;
Matches 480; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 4347 CAATTTCTACACAAAAGAGATGAGATTAAACAGTCAGGGTTGGGCTCTGCATTGGAGGTG 4406
Db 493 CAATTTCTACACAAAAGAGATGAGATTAAACAGTCAGGGTTGGGCTCTGCATTGGAGGTG 434
Qy 4407 CCCTATAAACACAGAAGAAAAATCTGAAAGACACAGGGGACGGGACAGACACAGACACAGAC 4466
Db 433 CCCTATAAACACAGAAGAAAAATCTGAAAGACACAGGGGACGGGACAGACACAGACACAGAC 374
Qy 4467 CCAGAGTCTCCAAAGCACAGAGTGGCAAAACAAAACCCGAGCTGAGCATCAGGACCTTGC 4526
Db 373 CCAGAGTCTCCAAAGCACAGAGTGGCAAAACAAAACCCGAGTTCAGCATCAGGACCTTGC 314
Qy 4527 CTCGAATTTGCTTTCCAGTATTACGGTGCCTCTCTCTGCCCCCTTTCCAGGGTATCTGT 4586
Db 313 CTCGAATTTGCTTTCCAGTATTACGGTGCCTCTCTCTGCCCCCTTTCCAGGGTATCTGT 254
Qy 4587 GGGTTGCCAGCTGGGAGGGCAACCATAGCCACACACAGATTTCTCTGAAAGTTTACA 4646
Db 253 GGGTTGCCAGCTGGGAGGGCAACCATAGCCACACACAGATTTCTCTGAAAGTTTACA 194
Qy 4647 ATGAGTAGCATTTTGGGGTGTAGGGTGGCAGCTCCCAAGGCCCTGCCCCAGCCCCA 4706
Db 193 ATGAGTAGCATTTTGGGGTGTAGGGTGGCAGCTCCCAAGGCCCTGCCCCAGCCCCA 134
Qy 4707 CCCACTCATGACTCTAAGTGTGTGTATTAATTTATTTATTTGGAGATGTTATTTATT 4766
Db 133 CCCACTCATGACTCTAAGTGTGTGTATTAATTTATTTATTTGGAGATGTTATTTATT 74
Qy 4767 AGATGATATTTATTGCGAATTTCTATTTCTGTATTAACAAATAAAATGCTTGGCCCCAGA 4826
Db 73 AGATGATATTTATTGCGAATTTCTATTTCTGTATTAACAAATAAAATGCTTGGCCCCAGA 14
Qy 4827 A 4827
Db 13 A 13

RESULT 10
BG437538
LOCUS 853 bp mRNA linear EST 14-MAR-2001
DEFINITION 602489277F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4621530 5',
mRNA sequence.
ACCESSION BG437538
VERSION BG437538.1 GI:13344044
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 853)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DP/Gazdar
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI384 row: i column: 19
High quality sequence stop: 761.
FEATURES
Location/Qualifiers
1..853
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:4621530"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:

```

EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 154 a 301 c 232 g 166 t

Query Match 9.3%; Score 471.4; DB 12; Length 853;  
Best Local Similarity 93.2%; Pred. No. 5.6e-97;  
Matches 592; Conservative 0; Mismatches 31; Indels 12; Gaps 9;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGGCTGGGG 3422  
DB 220 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGGCTGGGG 279  
QY 3423 CAGAGACTCTGCCAGGGGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAAAC 3482  
DB 280 CAGAGACTCTGCCAGGGGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAAAC 339  
QY 3483 TCGCGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTTACTTTCGCTGGCTCA 3542  
DB 340 TCGCGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTTACTTTCGCTGGCTCA 399  
QY 3543 ACCGTCAGGCTGCCACTGTGAGCTGGCGGAGCCTGGCCACTTCTGACACAGCCTCC 3602  
DB 400 ACCGTCAGGCTGCCACTGTGAGCTGGCGGAGCCTGGCCACTTCTGACACAGCCTCC 458  
QY 3603 AGGGCTCTGTGGGAGCACTTGGGGGCTCATGGCAGCTCTGGGCTACCCAC-TGCCCCAG 3661  
DB 459 AGGGCTCTGTGGGAGCACTTGGGGGCTCATGGCAGCTCTGGGCTACCCACTTGACCCAG 518  
QY 3662 CGCGTGCCTGGAGC-TGAACCCACTTGGACTCTCTGCGCTTGGCCACAGTGAATTCCTCCA 3720  
DB 519 CGCGTGCCTGGAGC-TGAACCCACTTGGACTCTCTGCGCTTGGCCACAGTGAATTCCTCCA 578  
QY 3721 GAAGATGGAGCACTTCTGCTGCTGAAGGAGCTGACAGCTGCTGCGCTGGCGCTCGGCCAA 3780  
DB 579 GAAGATGGAGCACTTCTGCTGCTGAAGGAGCTGAAAGACTGCTGCTGCTGGCGCTCGGCCAC 638  
QY 3781 GGACTTCAA-CGGGCTCAAGAAAGATGAGGCTCTCAG-CAGCTGCTGAGTCAACC---TG 3835  
DB 639 GACTTCAAGCGGCTCAAGAAAGATGAGGCTCTCAGGCTCCAGCCACTGAGTCAACCCATGC 698  
QY 3836 CACTTGGGGGCTCATGGCTTCTGACTTCTGACC-TTCTCTCTTCTGCTCCCTTCAAC 3894  
DB 699 ACTTGGAGGCTCATGGGTTCTGACTTCTGACCTTCTCTCTCTCTGCTCCCTTCAAC 758  
QY 3895 CCTGCTCCCACTTT--GTGAGAGCCAGCCTGTATGCCACACCTGTTGAGCCAGGAGAC 3952  
DB 759 CCTGTTCCCACTTTTGTGAGAGCAGGCTGTATGCCAACACCTGTTGAGCCAGGAGAC 818  
QY 3953 AGAA-GCTGTGAGCCTCTGGCCCTTTCTTCTGGACCG 3986  
DB 819 CGAAGCTGTGAGCCTCTGGCCCTTTTCTTCTGGAGC 853

RESULT 11  
BQ948158  
LOCUS BQ948158 887 bp mRNA linear EST 21-AUG-2002  
DEFINITION AGENCOURT\_8813192 NIH\_MGC\_101 Homo sapiens cDNA clone IMAGE:6428214  
5', mRNA sequence.

ACCESSION BQ948158  
VERSION BQ948158.1 GI:22363636  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 887)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC <http://mgc.nci.nih.gov/>.

# TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLC2614 row: h column: 07  
High quality sequence stop: 674.

# FEATURES source

1. .887  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:6428214"  
/clone\_lib="NIH\_MGC\_101"  
/tissue\_type="epidermoid carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pOTB7; Site 1: EcoRI; Site 2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 154 a 325 c 240 g 167 t 1 others  
ORIGIN

Query Match 9.2%; Score 470.4; DB 14; Length 887;  
Best Local Similarity 99.6%; Pred. No. 9.5e-97;  
Matches 482; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGGCTGGGG 3421  
DB 278 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGGCTGGGG 337  
QY 3422 CAGAGACTCTGCCAGGGGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAA 3481  
DB 338 CAGAGACTCTGCCAGGGGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAA 397  
QY 3482 CTGCGGCTGACCCAGAACTAGAGGCTTACAGCCACTTCTGTGTTACTTTCGCTGGCCTC 3541  
DB 398 CTGCGGCTGACCCAGAACTAGAGGCTTACAGCCACTTCTGTGTTACTTTCGCTGGCCTC 457  
QY 3542 AACCGTCAGGCTGCGACTGTGAGCTGCGCGCAGCCTGGGCCACTTCTGCAACAGCCTC 3601  
DB 458 AACCGTCAGGCTGCGACTGTGAGCTGCGCGCAGCCTGGGCCACTTCTGCAACAGCCTC 517  
QY 3602 CAGGGCTGTGGGAGCACTTGGCGGCTCATGGCAGCTCTGGGCTACCCACTGGCCCGAG 3661  
DB 518 CAGGGCTGTGGGAGCACTTGGCGGCTCATGGCAGCTCTGGGCTACCCACTGGCCCGAG 577  
QY 3662 CCGCTGCTGGGACTGAAACCCACTTGGACTCTGCGCCCTTGCACAGTGAATTCCTCCAG 3721  
DB 578 CCGCTGCTGGGACTGAAACCCACTTGGACTCTGCGCCCTTGCACAGTGAATTCCTCCAG 637  
QY 3722 AAGATGAGCACTTCTGGCTGCTGAAGGAGCTGACAGACTGGCTGTGGCGCTGGCCAG 3781  
DB 638 AAGATGAGCACTTCTGGCTGCTGAAGGAGCTGACAGACTGGCTGTGGCGCTGGCCAG 697  
QY 3782 GACTTCAACCCGGCTCAAGAAAGATGAGCCTCAGCAGCTGCAGTGCAGTCAACCTGACCTG 3841  
DB 698 GACTTCAACCCGGCTCAAGAAAGATGAGCCTCAGCAGCTGCAGTGCAGTCAACCTGACCTG 757  
QY 3842 GGGG 3845  
DB 758 GGGG 761

```

RESULT 12.
BF002167/c
LOCUS
DEFINITION
7h23c10.x1 NCI_CGAP_Col6 Homo sapiens cDNA clone IMAGE:3316818 3',
mRNA sequence.
ACCESSION
BF002167
VERSION
BF002167.1 GI:10702442
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 467)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 458.
Location/Qualifiers
1. .467
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3316818"
/clone_lib="NCI_CGAP_Col6"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/notes="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Col6 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 117 a 113 c 108 g 129 t
ORIGIN
Query Match 9.1%; Score 465.4; DB 12; Length 467;
Best Local Similarity 99.8%; Pred. No. 1.2e-95;
Matches 466; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4360 AAGAGATGAGATTAAACAGTGCAGGGTGGGTCTGATGGAGTGCCCTATAAACAG 4419
DB 467 AAGAGATGAGATTAAACAGTGCAGGGTGGGTATGATGGAGTGCCCTATAAACAG 408
QY 4420 AAGAGAAATACTGAAAGCAGGGGAGGAGACAGACAGACAGACCCAGGAGTCTCCA 4479
DB 407 AAGAGAAATACTGAAAGCAGGGGAGGAGACAGACAGACCCAGGAGTCTCCA 348
QY 4480 AAGCAGAGTGGCAAAACAAACCCGAGCTGAGCATCAGGACCTTGCTCGAATTGTCTT 4539
DB 347 AAGCAGAGTGGCAAAACAAACCCGAGCTGAGCATCAGGACCTTGCTCGAATTGTCTT 288
QY 4540 CCAGTATACGGTCCCTCTCTCTGCCCCCTTTCCAGGGTATCTGTGGTTCAGGCT 4599
DB 287 CCAGTATACGGTCCCTCTCTCTGCCCCCTTTCCAGGGTATCTGTGGTTCAGGCT 228
QY 4600 GGGGAGGGCAACCATAGCCACACAGGATTTCTGAAATTTACATGCAGTAGCATT 4659
DB 227 GGGGAGGGCAACCATAGCCACACAGGATTTCTGAAATTTACATGCAGTAGCATT 168

```

```

QY 4660 TTGGGGTGTAGGTGGCAGCTCCCAAGGCCCTGCCCCCAGCCGCCACCTCATGACT 4719
DB 167 TTGGGGTGTAGGTGGCAGCTCCCAAGGCCCTGCCCCCAGCCGCCACCTCATGACT 108
QY 4720 CTAAGTGTGTGATTAAATATTTATTTATTTGGAGATGTTATTTATTTAGATGATATTTAT 4779
DB 107 CTAAGTGTGTGATTAAATATTTATTTATTTGGAGATGTTATTTATTTAGATGATATTTAT 48
QY 4780 TGCAGAAATTTCTATTTCTGTATTACAAATATAAATGCTTGCCTCCAG 4826
DB 47 TGCAGAAATTTCTATTTCTGTATTACAAATATAAATGCTTGCCTCCAG 1
RESULT 13
BF513712/c
LOCUS
DEFINITION
UI-H-BMI-amy-c-07-0-UI.sl NCI_CGAP_Sub7 Homo sapiens cDNA clone
IMAGE:3071604 3', mRNA sequence.
ACCESSION
BF513712
VERSION
BF513712.1 GI:11598891
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 454)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=No.
Location/Qualifiers
1. .454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3071604"
/clone_lib="NCI_CGAP_Sub7"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub7
is a subtracted library derived from NCI_CGAP_Sub6. The
NCI_CGAP_Sub7 library had 12 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub6 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE CloneIDs
132376-132391, 145608-145675, 150052-150285);
NCI_CGAP_Kid3 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE CloneIDs 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
3851-3854 (IMAGE CloneIDs 1414920-1417991, 1520904-152439
); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
3733-3735 (IMAGE CloneIDs 1257096-1258631, 1469064-1470983
, 1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459,
2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255
, 1144584-1145351). (6% of the driver population), plus a
pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE
CloneIDs 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE
CloneIDs 2710536-2712455) (4% of the driver population
), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE
CloneIDs 2712456-2723591) (10% of the driver population),
plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE
CloneIDs 2723592-2729326) (40% of the driver population),

```

plus a pool of 4032 clones from NCI CGAP Sub6 (IMAGE Clones 2728969-2733190) (40% of the driver population). Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806. TAG\_SEQ=None found"]

BASE COUNT 102 a 74 c 166 g 112 t

Query Match 8.9%; Score 452.4; DB 12; Length 454;  
Best Local Similarity 99.8%; Pred. No. 1.1e-92;  
Matches 453; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 409 AAGTGTCCACTGAATCCTCAGCTTGTGCTCCTCACTGAAATCCAGAAAACATATTTCCTCC 468  
Db 454 AAGTGTCCACTGAATCCTCAGCTTGTGCTCCTCACTTAAATCCAGAAAACATATTTCCTCC 395  
QY 469 CACTGCCACATCCCTCTTACAGCACCCCAACCTCGCTCTGGACTCTCTGGTATCCTGG 528  
Db 394 CACTGCCACATCCCTCTTACAGCACCCCAACCTCGCTCTGGACTCTCTGGTATCCTGG 335  
QY 529 GATGTCCAAACTCTGAGTGCATGACCAACAGCCGCACTCTCAATGCACTCTCT 588  
Db 334 GATGTCCAAACTCTGAGTGCATGACCAACAGCCGCACTCTCAATGCACTCTCT 275  
QY 589 CCGTTCCTGTCGCCACCTTGAGGCTGATGAAAGCCCTCATGAGTCCAACTTTTCC 648  
Db 274 CCGTTCCTGTCGCCACCTTGAGGCTGATGAAAGCCCTCATGAGTCCAACTTTTCC 215  
QY 649 CCACCTAACCAAGAACGGGAGTGAACCTCCACACTGCGACCGTTCCTCGAGAGTGAGCA 708  
Db 214 CCACCTAACCAAGAACGGGAGTGAACCTCCACACTGCGACCGTTCCTCGAGAGTGAGCA 155  
QY 709 CTAAATCTCTCAATCTTAACCCACCTTACACTTCCACACTCAGAAATCACTCCTAG 768  
Db 154 CTAAATCTCTCAATCTTAACCCACCTTACACTTCCACACTCAGAAATCACTCCTAG 95  
QY 769 AATATACCCAAACTAAGCCCAATAGGAGCAGCCGCACTAGTGGTCTAACCCCTATACCT 828  
Db 94 AATATACCCAAACTAAGCCCAATAGGAGCAGCCGCACTAGTGGTCTAACCCCTATACCT 35  
QY 829 TGCTTCCTATGGGTGAGTCTGTTCTTGGGCGCG 862  
Db 34 TGCTTCCTATGGGTGAGTCTGTTCTTGGGCGCG 1

RESULT 14  
AW044541/c 457 bp mRNA linear EST 09-MAR-2000  
LOCUS  
DEFINITION  
IMAGE:2554187 3', mRNA sequence.  
ACCESSION  
AW044541  
VERSION  
AW044541.1 GI:5905070  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 457)  
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LINL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1753 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 449.  
Location/Qualifiers  
1..457

/organism="Homo sapiens"  
/db xref="taxon:9606"  
/clone="IMAGE:2554187"  
/clone\_lib="Soares NSF\_F8\_9W\_OT\_PA\_P\_S1"  
/lab\_host="DH10B"  
/notes="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;  
Equal amounts of plasmid DNA from five normalized  
libraries were mixed, and ss circles were made in vitro.  
Following HAP purification, this DNA was used as tracer in  
a subtractive hybridization reaction. The driver was  
PCR-amplified cDNAs from pools of 5,000 clones made from  
the same 5 libraries. The pools consisted of the following  
libraries and clones: Soares Nhsf pool 1:  
305384-310919, 323208-325895 Soares Nhsf pool 1:  
145032-147335, 147720-148103, 148872-149255, 15002 -  
150407, 151176-152327 Soares Nhsf pool 1:  
758280-760583, 772104-774407 Soares Nhsf pool 1:  
304776-306311, 320136-322823, 326280-326663 Soares Nhsf  
pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 114 a 109 c 111 g 123 t  
ORIGIN  
Query Match 8.9%; Score 450.4; DB 10; Length 457;  
Best Local Similarity 99.8%; Pred. No. 3.3e-92;  
Matches 451; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4376 CAGTGCAGGGTGGGCTCTGATTTGGAGTGGTCCCTATTAACAGAGAAATACTGAA 4435  
Db 456 CAGTGCAGGGTGGGCTCTGATTTGGAGTGGTCCCTATTAACAGAGAAATACTGAA 397  
QY 4436 AGCACAGGGGCGAGGACAGACAGACAGAGAGTCTCCAAAGCACAGAGTGGCAA 4495  
Db 396 AGCACAGGGGCGAGGACAGACAGACAGAGAGTCTCCAAAGCACAGAGTGGCAA 337  
QY 4496 ACAAAACCCGAGCTGAGCATCAGGACCTTGCTCGAATTTGCTTCCAGTATTACGGTGGC 4555  
Db 336 ACAAAACCCGAGCTGAGCATCAGGACCTTGCTCGAATTTGCTTCCAGTATTACGGTGGC 277  
QY 4556 TCTTCTGCGCCCTTTCCAGGGTATCTGTGGGTGGCAGGCTGGGAGGCAACCATTA 4615  
Db 276 TCTTCTGCGCCCTTTCCAGGGTATCTGTGGGTGGCAGGCTGGGAGGCAACCATTA 217  
QY 4616 GCCACACACAGAGATTTCTGAAAGTTTACAAATCAGTAGCATTTTGGGGTGTAGGGTGG 4675  
Db 216 GCCACACACAGAGATTTCTGAAAGTTTACAAATCAGTAGCATTTTGGGGTGTAGGGTGG 157  
QY 4676 CAGTCCCCAGGCGCTGCGCCCGCCAGCCCGCCAGCCCGCCAGCTCTAAAGTGTGTATT 4735  
Db 156 CAGTCCCCAGGCGCTGCGCCCGCCAGCCCGCCAGCTCTAAAGTGTGTATT 97  
QY 4736 AATATTATTATTGGAGATGTTATTATTAGATGATATTATTAGCAATTTCTATTTC 4795  
Db 96 AATATTATTATTGGAGATGTTATTATTAGATGATATTATTAGCAATTTCTATTTC 37  
QY 4796 TTGTATTAAACAATAAATAATGCTTCCGCCAGAA 4827  
Db 36 TTGTATTAAACAATAAATAATGCTTCCGCCAGAA 5

RESULT 15  
AI916548/c 462 bp mRNA linear EST 16-DEC-1999  
LOCUS  
DEFINITION  
IMAGE:2554187 3', mRNA sequence.  
ACCESSION  
AI916548  
VERSION  
AI916548.1 GI:5636403  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 462)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapsb-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Insert length: 600 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 445.

Search completed: February 1, 2003, 12:59:11  
Job time : 10371.6 secs

FEATURES  
source  
1..462  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2299261"  
/clone\_lib="NCI CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;  
Plasmid DNA from the normalized library NCI CGAP Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 114 a 108 c 108 g 132 t  
ORIGIN  
Query Match 8.8%; Score 446.2; DB 9; Length 462;  
Best Local Similarity 99.3%; Pred. No. 3e-91;  
Matches 448; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 4377 AGTCAGGGTTGGGCTCTGCATTGGAGTGGCCCTATAAACAGAGAGAAAATCTGAAA 4436  
Db 462 AGTCAGGGTTGGGCTCTGCATTGGAGTGGCCCTATAAACAGAGAGAAAATCTGAAA 403  
QY 4437 GCACAGGGGAGGACAGACAGACAGACAGACAGAGTCTCCAAAGCACAGAGTGGCAAA 4496  
Db 402 GCACAGGGGAGGACAGACAGACAGACAGAGTCTCCAAAGCACAGAGTGGCAAA 343  
QY 4497 CAAAACCCGAGCTGAGCATCAGGACCTTGCCCTCGAAATTTGCTTCAGTATTACGGTGCCT 4556  
Db 342 CAAAACCCGAGCTGAGCATCAGGACCTTGCCCTCGAAATTTGCTTCAGTATTACGGTGCCT 283  
QY 4557 CTTCTCTGCCCCCTTTCCAGGGTATCTGTGGTGGTCCAGGCTGGGGAGGGCAACCATAG 4616  
Db 282 CTTCTCTGCCCCCTTTCCAGGGTATCTGTGGTGGTCCAGGCTGGGGAGGGCAACCATAG 223  
QY 4617 CCACACACAGGATTTCTCTGAAAGTTTACATGAGTAGCATTTTGGGGTGTAGGGTGGC 4676  
Db 222 CCACACACAGGATTTCTCTGAAAGTTTACATGAGTAGCATTTTGGGGTGTAGGGTGGC 163  
QY 4677 AGCTCCCCAGGGCCCTGCCCCCAGCCCCCAGCCACTCATGACTCTAAGTGTGTGTATTA 4736  
Db 162 AGCTCCCCAGGGCCCTGCCCCCAGCCCCCAGCCACTCATGACTCTAAGTGTGTGTATTA 103  
QY 4737 ATATTTATTTATTTGGAGATGTTATTTATTTAGATGATATTTATTTGCAGAAATTTCTATCT 4796  
Db 102 ATATTTATTTATTTGGAGATGTTATTTATTTAGATGATATTTATTTGCAGAAATTTCTATCT 43  
QY 4797 TGTATTACAAATAAATGCTTGGCCCGAGAA 4827  
Db 42 TGTATTACAAATAAATGCTTGGCCCGAGAA 12





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 1295.47 Seconds  
(without alignments)  
8843.084 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 5087

Sequence: 1 aacctggagtgccctggc.....ccttggtaagtcttctca 5087

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-n-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5087	100.0	5087	21	AAA39482 Human NNT-1 DNA.
2	5087	100.0	5087	24	ABK11648 Human novel neurot
3	5076	99.8	5088	19	AAV47511 Human neurotrophic
4	5076	99.8	5088	19	AAV22653 Human genomic DNA
5	1464.4	28.8	1790	21	AAA88546 Human interleukin-
6	1463.4	28.8	1710	20	AAK16161 Human cardiophi
7	535.4	10.5	1008	22	AAK51548 Human polynucleoti
8	529.4	10.4	729	22	AAO4201 Human cardiophi
9	525.4	10.3	861	22	AAH74484 Nucleotide sequenc

10	523.4	10.3	797	19	AAV47510	Human neurotrophic
11	523.4	10.3	797	19	AAV22652	cDNA encoding huma
12	523.4	10.3	797	21	AAA39481	Human NNT-1 cDNA
13	523.4	10.3	797	24	ABK11647	Human cDNA encodin
14	485.4	9.7	968	22	ABA09140	Human cardiophi
15	495.4	9.7	968	22	AAK52532	Human polynucleoti
16	493.4	9.7	995	22	ABA71773	Human foetal liver
17	493.4	9.7	495	22	ABA37855	Probe #16321 for g
18	493.4	9.7	495	22	AAK20142	Human brain expres
19	493.4	9.7	495	22	AAK46202	Human bone marrow
20	493.4	9.7	495	22	AAI25564	Probe #15497 for g
21	493.4	9.7	495	22	AAI52108	Probe #20794 used
22	493.4	9.7	495	24	ABS20513	Human genome-deriv
23	492.4	9.7	768	22	AAH99772	Human protein enco
24	492	9.7	492	22	ABA59256	Human foetal liver
25	492	9.7	492	22	ABA27996	Probe #6462 for ge
26	492	9.7	492	22	AAK07469	Human brain expres
27	492	9.7	492	22	AAK33253	Human bone marrow
28	492	9.7	492	22	AAI16384	Probe #6317 for ge
29	492	9.7	492	22	AAI39047	Probe #7733 used t
30	492	9.7	492	24	ABS08085	Human genome-deriv
31	437.2	8.6	465	24	ABL81689	Human ovarian canc
32	429.6	8.4	819	19	AAV47512	Mouse neurotrophic
33	429.6	8.4	819	19	AAV22654	cDNA encoding muri
34	429.6	8.4	819	21	AAA39483	Murine NNT-1 cDNA.
35	429.6	8.4	819	24	ABK11649	Mouse cDNA encodin
36	429.4	8.4	432	24	ABL81658	Human ovarian canc
37	421.8	8.3	648	21	AAA88547	Mouse interleukin-
38	355.8	7.0	360	24	ABL79632	Human ovarian canc
39	330	6.5	342	20	AAK51546	Human secreted pro
40	329	6.5	340	21	AAI14792	Human secreted pro
41	324.6	6.4	396	20	AAI16162	Human cardiophi
42	250.6	4.9	269	24	ABL82404	Human ovarian canc
43	250.6	4.9	275	21	AAA42199	Human secreted exp
44	132.8	3.8	231	24	ABL79661	Human ovarian canc
45	165.2	3.2	283	22	AAH23145	Osteoarthritis tis

#### ALIGNMENTS

##### RESULT 1

AAA39482  
ID AAA39482 standard; DNA; 5087 BP.

AC AAA39482;

XX 24-AUG-2000 (first entry)

DT Human NNT-1 DNA.

XX NNT-1; human; neurotrophic factor; neuroprotective; treatment;  
XX anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;  
XX nervous system degeneration; Alzheimer's disease; Parkinson's disease;  
XX anyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
XX Huntington's disease; peripheral neuropathy; neural retina degeneration;  
XX retinopathy; immune disorder; hematopoietic disorder; ss.

OS Homo sapiens.

XX US6054294-A.

XX 25-APR-2000.

XX 12-DEC-1997; 97US-0988819.

XX 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.

XX Chang M;

XX WPI; 2000-338492/29.

New nucleic acids encoding neurotrophic factors useful for stimulating growth of motor or sympathetic neurons for treating neuron cell damage

\_

PT

XX

PS Claim 1b; Column 31-36; 42pp; English.

XX This invention describes a novel nucleic acid molecule (I) encoding a novel neurotrophic factor (NNT-1) (II) which has neurotropic, neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and opthalmological activity. (I) is useful for producing NNT-1 polypeptides which are useful for treating patients in whom various cells of the central, autonomic, or peripheral nervous system have degenerated and/or have been damaged by congenital disease, trauma, mechanical damage, surgery, stroke, ischemia, infection, metabolic disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1 proteins are used to treat diseases like Alzheimer's, Parkinson's, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's disease, peripheral neuropathy induced by diabetes or other metabolic disorders, and/or dystrophies or degeneration of the neural retina such as retinitis pigmentosa, drug-induced retinopathies, stationary forms of night blindness, progressive cone-rod degeneration, immune disorders and hematopoietic disorders. (I) is effective in treating neurological conditions and promotes neuron regeneration. Neural functions are effectively restored in patients suffering from various neurological disorders. This sequence encodes the human NNT-1 protein described in the method of the invention.

QY 1681 CCTTCCCTCTTAGTGGCCCCCCCCAATCCCCATCCTGGCCCCCAGGACTAGGCATGTG 1740  
DB 1681 CCTTCCCTCTTAGTGGCCCCCCCCAATCCCCATCCTGGCCCCCAGGACTAGGCATGTG 1740  
QY 1741 GGCAGGCTCGCACCCGCTTGGCCCAATGCGCCACTGGCTGCGAGCCGAGCCGCGGCC 1800  
DB 1741 GGCAGGCTCGCACCCGCTTGGCCCAATGCGCCACTGGCTGCGAGCCGAGCCGCGGCC 1800  
QY 1801 TCCCCCTGGGGCGGGGAAGTCTCCTCTGTGTTTACACCGTGTGTGTGTCTCTTGGCG 1860  
DB 1801 TCCCCCTGGGGCGGGGAAGTCTCCTCTGTGTTTACACCGTGTGTGTGTCTCTTGGCG 1860  
QY 1861 GCGGGGTTGGGTGGGGAAGAGGGGCCCCACCTCCATGCTGCGGTTCCAGCTCGCCTC 1920  
DB 1861 GCGGGGTTGGGTGGGGAAGAGGGGCCCCACCTCCATGCTGCGGTTCCAGCTCGCCTC 1920  
QY 1921 TGGCCCCAGACCTTGGGGCCCTGTGCTGTGACCCAGAGGSCCTCCCTCGCTGCTCT 1980  
DB 1921 TGGCCCCAGACCTTGGGGCCCTGTGCTGTGACCCAGAGGSCCTCCCTCGCTGCTCT 1980  
QY 1981 CCGATCTAGCTGGGCTCCTAGGGGGTCAATGGGGGAAGGGGACTGTAGGGAACCCAGG 2040  
DB 1981 CCGATCTAGCTGGGCTCCTAGGGGGTCAATGGGGGAAGGGGACTGTAGGGAACCCAGG 2040  
QY 2041 CAGTAGTGCAGGGGTTTAGGGTGGATGGAGTTATGCTGAAGATTTGGGGGTGG 2100  
DB 2041 CAGTAGTGCAGGGGTTTAGGGTGGATGGAGTTATGCTGAAGATTTGGGGGTGG 2100  
QY 2101 TCCAGAGGTGTTCAAGAGCCAGAGAGAGAGAGAGAGGGTGGAGGAGCCGAGGACCC 2160  
DB 2101 TCCAGAGGTGTTCAAGAGCCAGAGAGAGAGAGAGGGTGGAGGAGCCGAGGACCC 2160  
QY 2161 ATGGGGAACCGGCCCTCTTCCGTTTCTTCCATCCAGACCTACTCTGGAG 2220  
DB 2161 ATGGGGAACCGGCCCTCTTCCGTTTCTTCCATCCAGACCTACTCTGGAG 2220  
QY 2221 CCAGGGAAGAAAGGAGAGAGTGGGGGAGCTGCTCAGGCCAGAGTACACCG 2280  
DB 2221 CCAGGGAAGAAAGGAGAGAGTGGGGGAGCTGCTCAGGCCAGAGTACACCG 2280  
QY 2281 AGGAAATTAGTTGTCTGTGCTTGTGACGCTGTGAACCTCCCTGGGGCCCTTGCCTA 2340  
DB 2281 AGGAAATTAGTTGTCTGTGCTTGTGACGCTGTGAACCTCCCTGGGGCCCTTGCCTA 2340  
QY 2341 TCCAGGCTCTCCCTCTGCTTCTCCCTTCCAGTTATACATCTCCCTCATCCCTT 2400  
DB 2341 TCCAGGCTCTCCCTCTGCTTCTCCCTTCTTCCAGTTATACATCTCCCTCATCCCTT 2400  
QY 2401 TCCCTGGGCCCCAGCCGCTCCCGAGGGTGGAAAGGGCTTGCCCTTCCCTATACC 2460  
DB 2401 TCCCTGGGCCCCAGCCGCTCCCGAGGGTGGAAAGGGCTTGCCCTTCCCTATACC 2460  
QY 2461 ATGCTGTCTTCATAGCCTCTCCTGTCTTACTCATGAGCTGCTCCATTTCTCCTT 2520  
DB 2461 ATGCTGTCTTCATAGCCTCTCCTGTCTTACTCATGAGCTGCTCCATTTCTCCTT 2520  
QY 2521 CTGCAACCTGCTCTATCAGTGAACCCCTTCTTTCGAGTGTAGTAGTACCCGCTC 2580  
DB 2521 CTGCAACCTGCTCTATCAGTGAACCCCTTCTTTCGAGTGTAGTAGTACCCGCTC 2580  
QY 2581 TCCCCAGCCCTCAGCTGGTGGGCTGGGTGTGTCAGGGCAATGGGGCTGTGTTCCA 2640  
DB 2581 TCCCCAGCCCTCAGCTGGTGGGCTGGGTGTGTCAGGGCAATGGGGCTGTGTTCCA 2640  
QY 2641 ATGGGCCACTCTATCTCTCTTGTCTTGTGTCAGAAAAACCTTGTCTCACTCCACTG 2700  
DB 2641 ATGGGCCACTCTATCTCTCTTGTCTTGTGTCAGAAAAACCTTGTCTCACTCCACTG 2700  
QY 2701 CCTCTCTAGTTCGGACCCCTTCTCTCTGGGTTTCCCTGCGAAATTTCTCCAAAGGA 2760  
DB 2701 CCTCTCTAGTTCGGACCCCTTCTCTCTGGGTTTCTCTCTGGGTTTCCCTGCGAAATTTCTCCAAAGGA 2760  
QY 2761 GTGGTCTACACCCCTCTGCTCCACTTCTCTCTCCACCCCACTCACTTCTTAACCCCTGCAA 2820

DB 2761 GTGGTCTACACCCCTCTGCTCCACTCTCTCCACCACTCACTTCTTAACCCCTGCAA 2820  
QY 2821 TCTGGCTTCCAGCCCCCAGCAATGGTTCTTCCAAGTCTGTCAGGCACTCTCTTGCCTCAAG 2880  
DB 2821 TCTGGCTTCCAGCCCCCAGCAATGGTTCTTCCAAGTCTGTCAGGCACTCTCTTGCCTCAAG 2880  
QY 2881 CCGCAGAGTGTGTTGAAGGCTCAATTCCTTGTCTGTCTGTGTCAGGCACTCTGTGTGAG 2940  
DB 2881 CCGCAGAGTGTGTTGAAGGCTCAATTCCTTGTCTGTCTGTGTCAGGCACTCTGTGTGAG 2940  
QY 2941 CGTGTGCTCTTCTCGAACTCTCTTCTTGTGTCCTGCACTCTCTTGGGSCACCTTCTA 3000  
DB 2941 CGTGTGCTCTTCTCGAACTCTCTTCTTGTGTCCTTCTTGTGTCCTTGGGSCACCTTCTA 3000  
QY 3001 CTTCTCAGCTCTCTCAGGCTCTCTTCTCTGTCTGTCTGTGTCCTGCCCCACAGCGGCACTCTC 3060  
DB 3001 CTTCTCAGCTCTCTCAGGCTCTCTTCTCTGTCTGTCTGTGTCCTGCCCCACAGCGGCACTCTC 3060  
QY 3061 CCAAGGTTTGGCCCCCAGCCCAATCAGCAGCTCCTTCTGAGGCTCTTGTGCGTCTCTCTC 3120  
DB 3061 CCAAGGTTTGGCCCCCAGCCCAATCAGCAGCTCCTTCTGAGGCTCTTGTGCGTCTCTCTC 3120  
QY 3121 CTCTCTCTTCTTCTAGGCTCTCTTCTTGTGAGAGCTCACCACCGCCCACTGCTTCAACTGTC 3180  
DB 3121 CTCTCTCTTCTTCTAGGCTCTCTTCTTGTGAGAGCTCACCACCGCCCACTGCTTCAACTGTC 3180  
QY 3181 ACTGCTATCAAAATGATATCTTATTTGGAATACTCAGGAGGCCATGAACAAAGAGCC 3240  
DB 3181 ACTGCTATCAAAATGATATCTTATTTGGAATACTCAGGAGGCCATGAACAAAGAGCC 3240  
QY 3241 TAGCATGAGAGACAGGCGCAGTGTGAGGGGACACAAAATAGAACTTTGGGAGCAGGTA 3300  
DB 3241 TAGCATGAGAGACAGGCGCAGTGTGAGGGGACACAAAATAGAACTTTGGGAGCAGGTA 3300  
QY 3301 TCTCTTGTGTGTGAGCCAGCGGCTCTGCTCTCTCTTCTTCCCATCACCTCTCTCTTTTC 3360  
DB 3301 TCTCTTGTGTGTGAGCCAGCGGCTCTGCTCTCTCTTCTTCCCATCACCTCTCTCTTTTC 3360  
QY 3361 ACAGCTGAACCTACCTGGGCCCCCTTTTCAAGAGCAGACTTCAACCCCTCCCGCTGGG 3420  
DB 3361 ACAGCTGAACCTACCTGGGCCCCCTTTTCAAGAGCAGACTTCAACCCCTCCCGCTGGG 3420  
QY 3421 GGCAGAGACTCTGCCAGGCGCACCTGTGAGCTTACAGCCACCTTCTGTGTACTTTCGCTGGCT 3480  
DB 3421 GGCAGAGACTCTGCCAGGCGCACCTGTGAGCTTACAGCCACCTTCTGTGTACTTTCGCTGGCT 3480  
QY 3481 ACTGCGCTGAGCCAGAACTACAGGGCTTACAGCCACCTTCTGTGTACTTTCGCTGGCT 3540  
DB 3481 ACTGCGCTGAGCCAGAACTACAGGGCTTACAGCCACCTTCTGTGTACTTTCGCTGGCT 3540  
QY 3541 CAACCGTCAGGCTGCGACTGTGGGCTCATGAGCTTGGGCTTGGGCTTGGGCTTGGGCT 3600  
DB 3541 CAACCGTCAGGCTGCGACTGTGGGCTCATGAGCTTGGGCTTGGGCTTGGGCTTGGGCT 3600  
QY 3601 CCAGGGCTCTGCTGGGCGACTTGGGGCTCATGAGCTTGGGCTTGGGCTTGGGCTTGGGCT 3660  
DB 3601 CCAGGGCTCTGCTGGGCGACTTGGGGCTCATGAGCTTGGGCTTGGGCTTGGGCTTGGGCT 3660  
QY 3661 GCCGCTGCTGGGACTGAACCCACTTGGACTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 3720  
DB 3661 GCCGCTGCTGGGACTGAACCCACTTGGACTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 3720  
QY 3721 GAAGATGAGCAGCTTCTGGCTGTGAGGAGCTGAGACCTGGCTTGGGCTTGGGCTTGGGCT 3780  
DB 3721 GAAGATGAGCAGCTTCTGGCTGTGAGGAGCTGAGACCTGGCTTGGGCTTGGGCTTGGGCT 3780  
QY 3781 GGAATTCAACCGGCTCAAGAAAGATGAGCTTCCAGCAGCTGAGTCACTCCCTGCACT 3840  
DB 3781 GGAATTCAACCGGCTCAAGAAAGATGAGCTTCCAGCAGCTGAGTCACTCCCTGCACT 3840  
QY 3841 GGGGGCTCATGGCTTGTGACTTCTGACTTCTCTCTTCTGCTCTCTCTTCTTCAACCCCTGCT 3900

```
Db 3841 GGGGCTCATGGCTTCTGACTTCTGACCTTCTCCTCTTCGCTCCCTTCAAAACCTGCT 3900
Qy 3901 CCCATTGTGTGAGAGCCGCTGTATGCAACACACCTGTGTGAGCCAGGAGACAGAAGCTG 3960
Db 3901 CCCATTGTGTGAGAGCCGCTGTATGCAACACACCTGTGTGAGCCAGGAGACAGAAGCTG 3960
Qy 3961 TGAGCCTCTGGCCCTTCTGACCGGCTGGGGGTGTGATGGATCAGCCCTGTCTCTTC 4020
Db 3961 TGAGCCTCTGGCCCTTCTGACCGGCTGGGGGTGTGATGGATCAGCCCTGTCTCTTC 4020
Qy 4021 CCCACCTCCCAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTGTCTCTG 4080
Db 4021 CCCACCTCCCAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTGTCTCTG 4080
Qy 4081 TTTTCTACAGGAAGTCACTGCTGAGGAGGTGTGAAGTGGTTTCAAGTGGTTCAGGTTGGTGCAGAGGCGCT 4140
Db 4081 TTTTCTACAGGAAGTCACTGCTGAGGAGGTGTGAAGTGGTTTCAAGTGGTTCAGGTTGGTGCAGAGGCGCT 4140
Qy 4141 CATGGCTCTGCTTCTTGCTTACCACTTGGCCAGTGGCCACCCAGCCCTCAGGTGGCA 4200
Db 4141 CATGGCTCTGCTTCTTGCTTACCACTTGGCCAGTGGCCACCCAGCCCTCAGGTGGCA 4200
Qy 4201 CATCTGGAGGCGAGGGTGTGAGGGCCACCCACACATGCTTCTTGGGGTGAAGCCCT 4260
Db 4201 CATCTGGAGGCGAGGGTGTGAGGGCCACCCACACATGCTTCTTGGGGTGAAGCCCT 4260
Qy 4261 TTGGCTGCCCACTCTCTCTGGATGGGTGTGCTCCCTTATCCCAATCACTCTATACA 4320
Db 4261 TTGGCTGCCCACTCTCTCTGGATGGGTGTGCTCCCTTATCCCAATCACTCTATACA 4320
Qy 4321 TCCNATTCAGGAACAAACATGTGGCAATTTCTACACAAAGAGATGAGATTAACAGTG 4380
Db 4321 TCCNATTCAGGAACAAACATGTGGCAATTTCTACACAAAGAGATGAGATTAACAGTG 4380
Qy 4381 CAGGGTTGGGGTCTGCATTGGAGTGGCTTATAAACAGAGAGAAATACTGAAAGCAC 4440
Db 4381 CAGGGTTGGGGTCTGCATTGGAGTGGCTTATAAACAGAGAGAAATACTGAAAGCAC 4440
Qy 4441 AGGGGAGGACAGACAGACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAAA 4500
Db 4441 AGGGGAGGACAGACAGACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAAA 4500
Qy 4501 ACCCGAGCTGAGCATCAGGACCTTGCTCGAATGTCTTCAGTATTAACGGTGCCTTTC 4560
Db 4501 ACCCGAGCTGAGCATCAGGACCTTGCTCGAATGTCTTCAGTATTAACGGTGCCTTTC 4560
Qy 4561 TCTGCCCTTTTCCAGGGTATCTGGGGTTCAGGGTTCAGGGTGGGAGGCAACATAGCCAC 4620
Db 4561 TCTGCCCTTTTCCAGGGTATCTGGGGTTCAGGGTTCAGGGTGGGAGGCAACATAGCCAC 4620
Qy 4621 ACCACAGATTTCTGAAAGTTTACAATGCAATAGTAGCATTTTGGGGTGTAGGTGGCAGCT 4680
Db 4621 ACCACAGATTTCTGAAAGTTTACAATGCAATAGTAGCATTTTGGGGTGTAGGTGGCAGCT 4680
Qy 4681 CCCCAAGGCCCTGCCCCCAGCCACCCACTCATGACTCTAAGTGTGTGTATTATATAT 4740
Db 4681 CCCCAAGGCCCTGCCCCCAGCCACCCACTCATGACTCTAAGTGTGTGTATTATATAT 4740
Qy 4741 TTAATTTATTTGGAGATGTTATTTATAGATGATATTTATTCAGAAATTTCTATTCTTGA 4800
Db 4741 TTAATTTATTTGGAGATGTTATTTATAGATGATATTTATTCAGAAATTTCTATTCTTGA 4800
Qy 4801 TTAACAATAAATGCTTGGCCAGAACTTGTCTTTTGGCCAGCCCTCACCCCTCTCTGG 4860
Db 4801 TTAACAATAAATGCTTGGCCAGAACTTGTCTTTTGGCCAGCCCTCACCCCTCTCTGG 4860
Qy 4861 TGCTCATCAGACTCTTGCCACCCCTGGCTCCCACTTCCCTGTGTCTCTGTGTGGAGTGC 4920
Db 4861 TGCTCATCAGACTCTTGCCACCCCTGGCTCCCACTTCCCTGTGTCTCTGTGTGGAGTGC 4920
Qy 4921 ACAGAGCTCTGGAGAGGCGCTCTTCTCTCCCGACCTGGGGCGATGGGCGACCTCAGA 4980
Db 4921 ACAGAGCTCTGGAGAGGCGCTCTTCTCTCCCGACCTGGGGCGATGGGCGACCTCAGA 4980

Qy 4981 CTTTACCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCCACACAGCTTC 5040
Db 4981 CTTTACCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCCACACAGCTTC 5040
Qy 5041 TGTCCACCCAGGTTTCCCTCACCCACCCCTTGTAAAGTCTTCCTCA 5087
Db 5041 TGTCCACCCAGGTTTCCCTCACCCACCCCTTGTAAAGTCTTCCTCA 5087

RESULT 2
ABK11648
ID ABK11648 standard; DNA; 5087 BP.
XX
AC ABK11648;
XX
DT 05-JUN-2002 (first entry)
XX
Human novel neurotrophic factor NNT1, genomic DNA.
XX
Human; ds; gene; NNT1; neurotrophic factor; IGE-related disease;
Type I allergic disease; allergic rhinitis; eczema; dermatitis;
pollinosis; asthma; immune disease; cancer; arteriosclerosis;
vascular stenosis; rheumatoid arthritis; psoriatic arthritis;
inflammatory arthritis; osteoarthritis; lupus; diabetes; endometriosis;
autoimmune disease; multiple sclerosis; transplant rejection; reproductive disorder;
inflammatory bowel disease; infertility; miscarriage; preterm labour.
XX
Homo sapiens.
XX
Key Location/Qualifiers
misc_feature 137..138
/*tag= a
/ note= "Intervening, unsequenced region of more than 1KB"
XX
WO200215977-A2.
XX
28-FEB-2002.
XX
17-AUG-2001; 2001WO-US25906.
XX
18-AUG-2000; 2000US-226436P.
XX
16-AUG-2001; 2001US-0931704.
XX
(AMGE-) AMGEN INC.
XX
Senaldi G;
XX
WPI; 2002-280867/32.
XX
Treating Immunoglobulin E-related disease, modulating IgE levels in a
patient, preventing IGE-related disease and treating allergic diseases,
involves administering NNT-1 inhibitor to a patient
XX
Claim 2; Fig 2; 63pp; English.
XX
The invention relates to treating Immunoglobulin E (IGE)-related disease,
modulating IGE levels in a patient, preventing an IGE-related disease,
and treating allergic diseases, comprising administering a
therapeutically effective amount of novel neurotrophic factor (NNT)-1
inhibitor to a patient. Also included are a method of diagnosing an
IGE-related disease or susceptibility to an IGE-related disease, by
determining the presence or amount of expression of an NNT1 polypeptide
encoded by a NNT1 nucleotide sequence, its fragment or naturally
occurring variant, and diagnosing an IGE-related disease or
susceptibility of an IGE-related disease based on the presence or amount
of expression of the polypeptide and a pharmaceutical composition for use
in treating IGE-related disease, comprising the NNT1 inhibitor.
The NNT1 inhibitor is useful for preventing and treating IGE-related
disease, modulating IGE levels, and treating allergic diseases e.g.
Type I allergic disease, allergic rhinitis, eczema, dermatitis,
pollinosis, asthma, immune diseases and disorders, diseases involving
```

abnormal cell proliferation including cancer, arteriosclerosis and vascular remodeling, diseases and conditions relating to dysfunction of immune system including rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, osteoarthritis, inflammatory joint disease, autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory bowel disease, transplant rejection, and graft versus host disease, and reproductive diseases and disorders including infertility, miscarriage, preterm labour and delivery, and endometriosis. The present sequence is the genomic DNA for human NNT1.

SQ Sequence 5087 BP; 992 A; 1746 C; 1191 G; 1158 T; 0 other;

Query Match 100.0%; Score 5087; DB 24; Length 5087;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AAC	TGCGAGTGGCGCTGGCGGATGGGATATTAAAGCTTTCGCGGAGCGCGGCTCGCC	60
DB	1	AAC	TGCGAGTGGCGCTGGCGGATGGGATATTAAAGCTTTCGCGGAGCGCGGCTCGCC	60
QY	61	CTCC	CACTCCGCCAGGCTTCGGGAGAGCGCGCACCGCGCCGCGCCAGCGCCCA	120
DB	61	CTCC	CACTCCGCCAGGCTTCGGGAGAGCGCGCACCGCGCCGCGCCAGCGCCCA	120
QY	121	TGG	ACCTCCGAGCAGGTTGAAAACCCAAATAGCCCTGCTCTTTATAACATGACAAGCAG	180
DB	121	TGG	ACCTCCGAGCAGGTTGAAAACCCAAATAGCCCTGCTCTTTATAACATGACAAGCAG	180
QY	181	CGCC	CACTGTATACCTTAAACCGACCAAGTCA GAGCCCTCAACTCACTCTGCTGCTG	240
DB	181	CGCC	CACTGTATACCTTAAACCGACCAAGTCA GAGCCCTCAACTCACTCTGCTGCTG	240
QY	241	CAG	ACCTCAGCACATCTTGTGACTCAAAACCTCAACCGCACCTAAATCAACAAATCCCA	300
DB	241	CAG	ACCTCAGCACATCTTGTGACTCAAAACCTCAACCGCACCTAAATCAACAAATCCCA	300
QY	301	AGT	CTAAACTTAATCTGAAACTTTTAAAGTAAACCGAGTCTTTAAACCTTAACGCGCAAT	360
DB	301	AGT	CTAAACTTAATCTGAAACTTTTAAAGTAAACCGAGTCTTTAAACCTTAACGCGCAAT	360
QY	361	GCC	AATTATATCTACCTTAGCCAAACCCCTAACTGCTTTGCCAGTCAAAGTGTCCACTG	420
DB	361	GCC	AATTATATCTACCTTAGCCAAACCCCTAACTGCTTTGCCAGTCAAAGTGTCCACTG	420
QY	421	AAT	CTCACCCTTGGTCTCTACTGAAAATCCAGAAAAGCATATTTCCCACTGCCACAT	480
DB	421	AAT	CTCACCCTTGGTCTCTACTGAAAATCCAGAAAAGCATATTTCCCACTGCCACAT	480
QY	481	CCCT	CTTTACAGACCAACCCCTGGGCTCTGGACTCTGGTATCTCGGTATCTGGATGTCAAACT	540
DB	481	CCCT	CTTTACAGACCAACCCCTGGGCTCTGGACTCTGGTATCTCGGTATCTGGATGTCAAACT	540
QY	541	CTG	CAGTGCCATCAGCCAAACAAGCCGACTCGTCAAATGCACCTCTCTCCCTTCCTGTCC	600
DB	541	CTG	CAGTGCCATCAGCCAAACAAGCCGACTCGTCAAATGCACCTCTCTCCCTTCCTGTCC	600
QY	601	CC	ACCTTCGAGCTGATGGAAGGCTCATTTGAAGTCCAACTTTCCCGACCTTAACAC	660
DB	601	CC	ACCTTCGAGCTGATGGAAGGCTCATTTGAAGTCCAACTTTCCCGACCTTAACAC	660
QY	661	AAGA	ACGGGTGAACCTCCACACTGCCACCGTTTCCTGAGAGTGAGCACTAAATCTCCTT	720
DB	661	AAGA	ACGGGTGAACCTCCACACTGCCACCGTTTCCTGAGAGTGAGCACTAAATCTCCTT	720
QY	721	CA	ATCTAAACCCCAACCTTACACTTTCCCACTCAGGAATCACTCTAGGAATATACCCAA	780
DB	721	CA	ATCTAAACCCCAACCTTACACTTTCCCACTCAGGAATCACTCTAGGAATATACCCAA	780
QY	781	ACT	AAGCCCATAAAGCAGCGCCGACCTAGTGGTCTAAACCTTATACCTTCTTCTATGG	840
DB	781	ACT	AAGCCCATAAAGCAGCGCCGACCTAGTGGTCTAAACCTTATACCTTCTTCTATGG	840
QY	841	GT	GAGTCTGTCTTGGCGCGCGCTCTCTCCTGCTTCTCTCTCTTGAAGCTGACTGTCTC	900

Db 1921 TGCCCCAGACCTGGGGCCCTGCTGCTCTGAGACCCAGGGGCTCCCTTCGCTGCTCTC 1980  
Qy 1981 CCCATCTAGCTGGGCTCCTTAGGGGGTCTATGGGGGAAGGGACTGTAGGGAACCCAGG 2040  
Db 1981 CCCATCTAGCTGGGCTCCTTAGGGGGTCTATGGGGGAAGGGACTGTAGGGAACCCAGG 2040  
Qy 2041 CAGTAGTGCAGGGGTTTAGGGTGTGGATGGAGTTATGCTGTAAAGATTTCGGGGTGG 2100  
Db 2041 CAGTAGTGCAGGGGTTTAGGGTGTGGATGGAGTTATGCTGTAAAGATTTCGGGGTGG 2100  
Qy 2101 TCCAGAGGTGTTTCAGAGACCCAGAGAGAGAAAGAGAGGGTTGGAGGACCCAGGCACC 2160  
Db 2101 TCCAGAGGTGTTTCAGAGACCCAGAGAGAGAAAGAGAGGGTTGGAGGACCCAGGCACC 2160  
Qy 2161 ATGGGGAACCGGCCCTCTTCCCGTGTCTCTTCCATCCAGACCCCTACTCTGGAG 2220  
Db 2161 ATGGGGAACCGGCCCTCTTCCCGTGTCTCTTCCATCCAGACCCCTACTCTGGAG 2220  
Qy 2221 CCAGGGAAGAAAGGAAGAGTGGCGGGGAGCTGGCTCCAGCCCCAGGATACACCG 2280  
Db 2221 CCAGGGAAGAAAGGAAGAGTGGCGGGGAGCTGGCTCCAGCCCCAGGATACACCG 2280  
Qy 2281 AGGAAATTAGTTGTTCTCTGTCTGTCTGTGTCAGCGTGTGAACCTCCCGCTGGGCCCTTGCTTA 2340  
Db 2281 AGGAAATTAGTTGTTCTCTGTCTGTCTGTGTCAGCGTGTGAACCTCCCGCTGGGCCCTTGCTTA 2340  
Qy 2341 TCCAGGCTCTCCCTTGTCT 2400  
Db 2341 TCCAGGCTCTCCCTTGTCT 2400  
Qy 2401 TCCCTGGGCCAGCGCTCCCGAGGGTTGGAAAGGGCTCTGCCCTCTCTCCCTATACC 2460  
Db 2401 TCCCTGGGCCAGCGCTCCCGAGGGTTGGAAAGGGCTCTGCCCTCTCTCCCTATACC 2460  
Qy 2461 ATGCTGTCTTCCATAGCCT 2520  
Db 2461 ATGCTGTCTTCCATAGCCT 2520  
Qy 2521 CTGCAACCTGCT 2580  
Db 2521 CTGCAACCTGCT 2580  
Qy 2581 TCCCGAGCCCTCAGCTGTGGGCTGGGTGTGTGAGCGCAATGGGGCTCTGGTTCCA 2640  
Db 2581 TCCCGAGCCCTCAGCTGTGGGCTGGGTGTGTGAGCGCAATGGGGCTCTGGTTCCA 2640  
Qy 2641 ATGGGCACTCTCATCT 2700  
Db 2641 ATGGGCACTCTCATCT 2700  
Qy 2701 CCTCTCTAGTTCCGACCCCTTTTCT 2760  
Db 2701 CCTCTCTAGTTCCGACCCCTTTTCT 2760  
Qy 2761 GTGGTCTACACCT 2820  
Db 2761 GTGGTCTACACCT 2820  
Qy 2821 TCTGGCTTCCAGGCCCCAGCAATGGTTCT 2880  
Db 2821 TCTGGCTTCCAGGCCCCAGCAATGGTTCT 2880  
Qy 2881 CCCGACAGTCTTTTGAAGGCTCATCT 2940  
Db 2881 CCCGACAGTCTTTTGAAGGCTCATCT 2940  
Qy 2941 CGCTGTCT 3000  
Db 2941 CGCTGTCT 3000  
Qy 3001 CCTCTCTAGCT 3060  
Db 3001 CCTCTCTAGCT 3060

Qy 3061 CCAAGGTTTCCCAACCCAGGCAATCAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3120  
Db 3061 CCAAGGTTTCCCAACCCAGGCAATCAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3120  
Qy 3121 CT 3180  
Db 3121 CT 3180  
Qy 3181 ACCTGCAATCAAAATGATATCTTTATTTGAAAACTCAGGAGGCCATCAAAAGAGCC 3240  
Db 3181 ACCTGCAATCAAAATGATATCTTTATTTGAAAACTCAGGAGGCCATCAAAAGAGCC 3240  
Qy 3241 TAGCATGGAGACAGGGCCAGTGTCTAGGGGACACAAAAATAGAACTTTGGGAGCAGGTA 3300  
Db 3241 TAGCATGGAGACAGGGCCAGTGTCTAGGGGACACAAAAATAGAACTTTGGGAGCAGGTA 3300  
Qy 3301 TCTCTTTGGTGTGAGCAGGGCTCTGCT 3360  
Db 3301 TCTCTTTGGTGTGAGCAGGGCTCTGCT 3360  
Qy 3361 ACAGCTGAACTTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGG 3420  
Db 3361 ACAGCTGAACTTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGG 3420  
Qy 3421 GGCAAGAGACTCTGCCCAAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAA 3480  
Db 3421 GGCAAGAGACTCTGCCCAAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAA 3480  
Qy 3481 ACTCGGCTGTGAGCCAGAACTACGAGGCCCTACAGCCACTTCTGTGTACTTCTGTGGCTCT 3540  
Db 3481 ACTCGGCTGTGAGCCAGAACTACGAGGCCCTACAGCCACTTCTGTGTACTTCTGTGGCTCT 3540  
Qy 3541 CAAACCTGAGCTGCCACTGCTGAGCTGCGCGAGCTGGCCACTTCTGACACCAAGCCT 3600  
Db 3541 CAAACCTGAGCTGCCACTGCTGAGCTGCGCGAGCTGGCCACTTCTGACACCAAGCCT 3600  
Qy 3601 CCAGGCTCTGTGGGAGCATTGGGGCTCATGCGAGCTTGGGTATCCCACTGCCCA 3660  
Db 3601 CCAGGCTCTGTGGGAGCATTGGGGCTCATGCGAGCTTGGGTATCCCACTGCCCA 3660  
Qy 3661 GCCGTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCACAGTACTTCTCTCCA 3720  
Db 3661 GCCGTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCACAGTACTTCTCTCCA 3720  
Qy 3721 GAAGATGAGCAGCTTCTGGCTGTGAAGAGCTGAGAGCTGAGACTTGGCTGTGGGCTCGGCCAA 3780  
Db 3721 GAAGATGAGCAGCTTCTGGCTGTGAAGAGCTGAGAGCTGAGACTTGGCTGTGGGCTCGGCCAA 3780  
Qy 3781 GGACTTCAACCGGCTCAAGAAAGATGAGCCTCCAGCAGCTGAGTCAACCTGACCT 3840  
Db 3781 GGACTTCAACCGGCTCAAGAAAGATGAGCCTCCAGCAGCTGAGTCAACCTGACCT 3840  
Qy 3841 GGGGCTCATGGCTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900  
Db 3841 GGGGCTCATGGCTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900  
Qy 3901 CCCACTTTGTAGAGCCAGCCCTGTATGCCAACACTGTGTGAGCCAGGAGACAGAGCTG 3960  
Db 3901 CCCACTTTGTAGAGCCAGCCCTGTATGCCAACACTGTGTGAGCCAGGAGACAGAGCTG 3960  
Qy 3961 TGAGCTCTGGCCCTTTCTTGACCCGGCTGGGCTGTGTATGCGATCAGCCCTGTCTCTCTCT 4020  
Db 3961 TGAGCTCTGGCCCTTTCTTGACCCGGCTGGGCTGTGTATGCGATCAGCCCTGTCTCTCTCTCT 4020  
Qy 4021 CCCACTCCAAAGCTTACCGAGCTGGGAGGAGTACAGTAGGCCCTGTCTCTCTCTCTCTCTCTCT 4080  
Db 4021 CCCACTCCAAAGCTTACCGAGCTGGGAGGAGTACAGTAGGCCCTGTCTCTCTCTCTCTCTCTCT 4080  
Qy 4081 TTTCTACAGGAAGTCAATGCTCCAGGAGTGTGAAGTGGTTTCTAGGTTGGTGCAGAGGCGCT 4140  
Db 4081 TTTCTACAGGAAGTCAATGCTCCAGGAGTGTGAAGTGGTTTCTAGGTTGGTGCAGAGGCGCT 4140





CC inflammation. NNT-1 is also able to boost immunoreactivity and  
CC antibody production following vaccination, and, since it inhibits  
CC tumour necrosis factor production, it may also be useful for  
CC treating sepsis. NNT-1 nucleic acid fragments are also used as  
CC hybridisation probes in diagnostic assays. In addition, cells that  
CC have been engineered to express NNT-1 can be implanted, or nucleic  
CC acids are delivered in gene therapy vectors.

XX Sequence 5088 BP; 992 A; 1746 C; 1191 G; 1158 T; 1 other;

Query Match 99.8%; Score 5076; DB 19; Length 5088;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5087; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy 1 AACCTGCGAGTGGGCTGGCGATGGGATTATTAAGCTTCGCCGAGCGCGGCTCGCC 60
Db 1 AACCTGCGAGTGGGCTGGCGATGGGATTATTAAGCTTCGCCGAGCGCGGCTCGCC 60

Qy 61 CTCCTCACTCCGCGAGCTCCGGAGAGGAGCGCACCCCGCGCGCCAGCCCCAGCCCCA 120
Db 61 CTCCTCACTCCGCGAGCTCCGGAGAGGAGCGCACCCCGCGCGCCAGCCCCAGCCCCA 120

Qy 121 TGGACCTCGAGCAGGTTGAAGAACCCAAACTAGCCCTGCTCTTTCAATGAACAAGCAG 180
Db 121 TGGACCTCGAGCAGGTTGAAGAACCCAAACTAGCCCTGCTCTTTCAATGAACAAGCAG 180

Qy 181 CGCCCACTGATACCTAAACGACCAAGTCAAGCGCTCCCACTCAGCTCTGCCCTGCC 240
Db 181 CGCCCACTGATACCTAAACGACCAAGTCAAGCGCTCCCACTCAGCTCTGCCCTGCC 240

Qy 241 CAGACCTCAACACATCCTTTG-TGGACTCAAACTCAACCGCACTAAATCAACCAATCCC 299
Db 241 CAGACCTCAACACATCCTTTG-TGGACTCAAACTCAACCGCACTAAATCAACCAATCCC 300

Qy 300 AAGTCTAAACTAATCTGAACCTTTAAAGTAAACCCAGTCTTTAAACCTAAGGCCAA 359
Db 301 AAGTCTAAACTAATCTGAACCTTTAAAGTAAACCCAGTCTTTAAACCTAAGGCCAA 360

Qy 360 TGGCAATATATACCTAGCAACCAACCTTAAGTCTGCTTGGCAGTCCCAAGTGTCCACT 419
Db 361 TGGCAATATATACCTAGCAACCAACCTTAAGTCTGCTTGGCAGTCCCAAGTGTCCACT 420

Qy 420 GAATCTCACTTGGTCTCACTGAAATCCAGAAAGCATATTTCCCACTGCCACCA 479
Db 421 GAATCTCACTTGGTCTCACTGAAATCCAGAAAGCATATTTCCCACTGCCACCA 480

Qy 480 TCCCTCTTACAGCAACCAACCTGGCTCTGGACTCTGCTGATTCCTGGGATGTCGAAC 539
Db 481 TCCCTCTTACAGCAACCAACCTGGCTCTGGACTCTGCTGATTCCTGGGATGTCGAAC 540

Qy 540 TCTGAGTGGCATCAGCCCAACAGCCGACTCGTCAATGACCTCTCCCTTCCTGTC 599
Db 541 TCTGAGTGGCATCAGCCCAACAGCCGACTCGTCAATGACCTCTCCCTTCCTGTC 600

Qy 600 CCCACCTTGCAGGCTGATGAAAGGCTCATTTGAAGTCAAACTTTTCCCACTTAACAC 659
Db 601 CCCACCTTGCAGGCTGATGAAAGGCTCATTTGAAGTCAAACTTTTCCCACTTAACAC 660

Qy 660 CAAGAACGGGGTGAACCTCCACACTGCGACCGTTCCTGAGAGTGAGCACTAAATCTCCT 719
Db 661 CAAGAACGGGGTGAACCTCCACACTGCGACCGTTCCTGAGAGTGAGCACTAAATCTCCT 720

Qy 720 TCAATCTAACCCCACTTACCTTCCACACTCAGGAATCACTCTAGATATACCCAA 779
Db 721 TCAATCTAACCCCACTTACCTTCCACACTCAGGAATCACTCTAGATATACCCAA 780

Qy 780 AACTAAGCCCAATAAGCAGCCGACCTAGTGGTCTAAGCCCTATACCTTCTTCTATG 839
Db 781 AACTAAGCCCAATAAGCAGCCGACCTAGTGGTCTAAGCCCTATACCTTCTTCTATG 840

Qy 840 GGTGAGTCTGTTTGGGGCGGCTCTCTCTGCTTCTCTGCTTCTAGAGTGTGCT 899
Db 841 GGTGAGTCTGTTTGGGGCGGCTCTCTCTGCTTCTCTGCTTCTAGAGTGTGCT 900
```

```
Qy 900 CAGCCTGCCAGCTCTGACATGTGCTGTCTCCCACTCTGACTCCCTCAAGCTGCACTG 959
Db 901 CAGCCTGCCAGCTCTGACATGTGCTGTCTCCCACTCTGACTCCCTCAAGCTGCACTG 960

Qy 960 GGAAGTGAAGAGTGGAGGAGTGGGTGACAACTGGAACACAGCAGGCTGCACTGCG 1019
Db 961 GGAAGTGAAGAGTGGAGGAGTGGGTGACAACTGGAACACAGCAGGCTGCACTGCG 1020

Qy 1020 TCCTTAGGCTGGCCCGCTCCCTCCATGTACACATATACATGTGGCACACACAGT 1079
Db 1021 TCCTTAGGCTGGCCCGCTCCCTCCATGTACACATATACATGTGGCACACACAGT 1080

Qy 1080 GGCACATATGCCAAAGACTCTCTCAGCTGACACACAGATCCATTTCTAAGTATCTACTGA 1139
Db 1081 GGCACATATGCCAAAGACTCTCTCAGCTGACACACAGATCCATTTCTAAGTATCTACTGA 1140

Qy 1140 TAGACATCATGGTGGCAAGTCTCATCTCTCAACATACACATGCCTCTCTTCTCTCC 1199
Db 1141 TAGACATCATGGTGGCAAGTCTCATCTCTCAACATACACATGCCTCTCTTCTCTCC 1200

Qy 1200 CGTCTTGGCAGGAGTGTTCCTCCCTCCATCCCTCTGCTCCCTCCATCTGGTGTCCCACT 1259
Db 1201 CGTCTTGGCAGGAGTGTTCCTCCCTCCATCCCTCTGCTCCCTCCATCTGGTGTCCCACT 1260

Qy 1260 CTCACCCCCCAACCCAGCCCAAGTGGGAGACAGACCTGAGGGGCTGCAGCTGCTTCCC 1319
Db 1261 CTCACCCCCCAACCCAGCCCAAGTGGGAGACAGACCTGAGGGGCTGCAGCTGCTTCCC 1320

Qy 1320 CGTGTGGGCGGGCGGGCTCATGCTCTGCTGCTCATCTGCTCCCAACAGGGACTCGTGG 1379
Db 1321 CGTGTGGGCGGGCGGGCTCATGCTCTGCTGCTCATCTGCTCCCAACAGGGACTCGTGG 1380

Qy 1380 GGATGTTAGCGTGTGCTGTGACCGTGTCTGCGACCTCTGCTGAGTGCAGCTCTCAATC 1439
Db 1381 GGATGTTAGCGTGTGCTGTGACCGTGTCTGCGACCTCTGCTGAGTGCAGCTCTCAATC 1440

Qy 1440 GCACAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCCGCTACC 1499
Db 1441 GCACAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCCGCTACC 1500

Qy 1500 TGGAGCACTCGAGCTTGGCTGGAGCTATGCTGAGTATCCAGCTAGGAATCTGG 1559
Db 1501 TGGAGCACTCGAGCTTGGCTGGAGCTATGCTGAGTATCCAGCTAGGAATCTGG 1560

Qy 1560 GAGTTGGGAGAGTGGAGAGTGGGAAAGACAGTCTTAAACCGTGGAGGGTTCTGGTAA 1619
Db 1561 GAGTTGGGAGAGTGGAGAGTGGGAAAGACAGTCTTAAACCGTGGAGGGTTCTGGTAA 1620

Qy 1620 ATGATGGGCTGAGGAGGGCTCTTTGGCTTCCACACAGTCCCTCTGCTGCTATCTCCT 1679
Db 1621 ATGATGGGCTGAGGAGGGCTCTTTGGCTTCCACACAGTCCCTCTGCTGCTATCTCCT 1680

Qy 1680 GCCCTTCCCTCTTAGTGGGCGGGCTCTTCCCACTCCCTGCGCCAGGACTAGGATGT 1739
Db 1681 GCCCTTCCCTCTTAGTGGGCGGGCTCTTCCCACTCCCTGCGCCAGGACTAGGATGT 1740

Qy 1740 GGGCAGGCTCGACCCCGCTTGGCCCATTTGCCCACTGGCTGCCAGCCAGCCGCGCGC 1799
Db 1741 GGGCAGGCTCGACCCCGCTTGGCCCATTTGCCCACTGGCTGCCAGCCAGCCGCGCGC 1800

Qy 1800 CTCCCTCTGGGGCGGGGAAAGTCTCTCTGTTTACACCGTGTGTTGTTCTCTTGGCG 1859
Db 1801 CTCCCTCTGGGGCGGGGAAAGTCTCTCTGTTTACACCGTGTGTTGTTCTCTTGGCG 1860

Qy 1860 GGGCGGGGTTGGGTGGGACAGAGGGGCGCCCTCCCTCCATGCTGCTTCCAGTCCGCT 1919
Db 1861 GGGCGGGGTTGGGTGGGACAGAGGGGCGCCCTCCCTCCATGCTGCTTCCAGTCCGCT 1920

Qy 1920 CTGCCCCCAGACCTGGGGCGGCTGCTGCTGAGACCCAGGGGCGCTCCCTTCCGCTC 1979
Db 1921 CTGCCCCCAGACCTGGGGCGGCTGCTGCTGAGACCCAGGGGCGCTCCCTTCCGCTC 1980
```





Db 4141 TCATGGCTCTGCTTCTTGGCTTACACATCTGGCCAGTGGCCACCCAGCCCTCAGGTGGC 4200  
Qy 4200 ACATCTGGAGGCGAGGGTTGAGGGGCCACACCAACATGCTTCTTGGGGTGAAGCCC 4259  
Db 4201 ACATCTGGAGGCGAGGGTTGAGGGGCCACACCAACATGCTTCTTGGGGTGAAGCCC 4260  
Qy 4260 TTTGGTGGCCCACTCTCTTGGATGGGTGTTGCTCCCTTATCCCAAACTCATATAC 4319  
Db 4261 TTTGGTGGCCCACTCTCTTGGATGGGTGTTGCTCCCTTATCCCAAACTCATATAC 4320  
Qy 4320 ATCCAAATTCAGGAAACAAACATGGTGGCAATCTACACAAAGAGATGAGATTAACAGT 4379  
Db 4321 ATCCAAATTCAGGAAACAAACATGGTGGCAATCTACACAAAGAGATGAGATTAACAGT 4380  
Qy 4380 GCAGGGTTGGGTCTGCATTTGGAGTGGCTTATAACACAGAGAAAATCTGAAGCA 4439  
Db 4381 GCAGGGTTGGGTCTGCATTTGGAGTGGCTTATAACACAGAGAAAATCTGAAGCA 4440  
Qy 4440 CAGGGCAGGACAGACACAGACAGACAGAGAGTCTCCAAAGCAGAGTGGCAACAA 4499  
Db 4441 CAGGGCAGGACAGACACAGACAGAGAGTCTCCAAAGCAGAGTGGCAACAA 4500  
Qy 4500 AACCCGAGCTGAGCATCAGACCTTGGCTCGAATTTGCTTCCAGTATTACGGTGCCTTT 4559  
Db 4501 AACCCGAGCTGAGCATCAGACCTTGGCTCGAATTTGCTTCCAGTATTACGGTGCCTTT 4560  
Qy 4560 CTCTGCCCTTTCACAGGTATCTGTTGGTGGCAGCTGGGAGGACCACTAGCCA 4619  
Db 4561 CTCTGCCCTTTCACAGGTATCTGTTGGTGGCAGCTGGGAGGACCACTAGCCA 4620  
Qy 4620 CACACAGGATTTCTGAAATTTACATGCAATGAGTATTTGGGTGTAGGGTGGCAGC 4679  
Db 4621 CACACAGGATTTCTGAAATTTACATGCAATGAGTATTTGGGTGTAGGGTGGCAGC 4680  
Qy 4680 TCCCAAGGCGCTGCCCCCAGCCCCCAGCTCATGACTTAAGTGTGTTGTAATA 4739  
Db 4681 TCCCAAGGCGCTGCCCCCAGCCCCCAGCTCATGACTTAAGTGTGTTGTAATA 4740  
Qy 4740 TTTATTTATTTGGAGATTTATTTATTTAGATGATATTTATTTGCAGATTTCTATTCTGT 4799  
Db 4741 TTTATTTATTTGGAGATTTATTTATTTAGATGATATTTATTTGCAGATTTCTATTCTGT 4800  
Qy 4800 ATTAACAAATAAATGCTTGGCCAGAACTTAGTCTCTTTTGGCCAGCCTCACCCCTCTG 4859  
Db 4801 ATTAACAAATAAATGCTTGGCCAGAACTTAGTCTCTTTTGGCCAGCCTCACCCCTCTG 4860  
Qy 4860 GTGCTCATCAGACTCTTGGCCACCCCTGGCTCCCACTCCCTTGGTGGAGCTG 4919  
Db 4861 GTGCTCATCAGACTCTTGGCCACCCCTGGCTCCCACTCCCTTGGTGGAGCTG 4920  
Qy 4920 CACAGAGCTCTGGAAGAGGCGCTTCTTCCCTCCCGCACTGGGGCGATGGGCGCACTCAG 4979  
Db 4921 CACAGAGCTCTGGAAGAGGCGCTTCTTCTCCCGCACTGGGGCGATGGGCGCACTCAG 4980  
Qy 4980 ACTTACCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCTGCCACACAGCTT 5039  
Db 4981 ACTTACCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCTGCCACACAGCTT 5040  
Qy 5040 CTGTCCACCCAGGTTTCCCTCACCCGACCTTGTGTAAGTCTTCTCA 5087  
Db 5041 CTGTCCACCCAGGTTTCCCTCACCCGACCTTGTGTAAGTCTTCTCA 5088

RESULT 4  
AAV22653  
ID AAV22653 standard; DNA; 5088 BP.  
XX  
AC AAV22653;  
XX  
DT 13-JUL-1998 (first entry)  
XX  
DE Human genomic DNA encoding neurotrophic factor NNT-1.  
XX

KW Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;  
KW treatment; neurological disease; degeneration; Parkinson's disease;  
KW anyotropic lateral sclerosis; ALS; Alzheimer's disease; stroke; ss.  
OS Homo sapiens.  
XX  
PH Key Location/Qualifiers  
FT misc\_feature 138  
FT /\*tag= a  
FT /note= "represents intervening unsequenced region  
FT of 1 Kb"  
XX  
XX US5741772-A.  
XX  
XX 21-APR-1998.  
XX  
XX 03-FEB-1997; 97US-0792019.  
XX  
XX 03-FEB-1997; 97US-0792019.  
XX  
XX (AMGE-) AMGEN INC.  
XX  
XX Chang M;  
XX  
XX WPI; 1998-260526/23.  
XX  
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -  
XX useful for stimulating growth of motor and sympathetic neurons  
XX  
XX Disclosure; Fig 2; 41pp; English.  
XX  
XX The present sequence encodes a human neurotrophic factor, designated  
XX NNT-1, which is capable of stimulating growth of motor or sympathetic  
XX neurons. The NNT-1 protein is useful in the treatment of neurological  
XX diseases characterised by the degeneration and death of particular  
XX classes of neurons. These diseases specifically include Parkinson's  
XX disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,  
XX stroke and various degenerative disorders affecting vision.  
XX  
XX Sequence 5088 BP; 992 A; 1746 C; 1191 G; 1158 T; 1 other;  
XX

Query Match 99.8%; Score 5076; DB 19; Length 5088;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5087; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
Qy 1 AACCTGCGAGTGGCGCTGGCGATGGATTTAAAGCTTTCGCGGAGCGCGCTCGCC 60  
Db 1 AACCTGCGAGTGGCGCTGGCGATGGATTTAAAGCTTTCGCGGAGCGCGCTCGCC 60  
Qy 61 CTCCCACTCCGAGCTCGGAGAGGAGCGCGACCGCGCGCCAGCCAGCCCAAGCCCA 120  
Db 61 CTCCCACTCCGAGCTCGGAGAGGAGCGCGACCGCGCGCCAGCCAGCCCAAGCCCA 120  
Qy 121 TGGACCTCGAGCAGT-TGMAAACCAACTAGCCTCTCTTCAATCAATCAAGCA 179  
Db 121 TGGACCTCGAGCAGTNTGMAAACCAACTAGCCTCTCTTCAATCAATCAAGCA 180  
Qy 180 GGGCCCCATCTGTATACCTAAACGACCACTCAGCCCTCCAACTCACCCTTGCCTGC 239  
Db 181 GGGCCCCATCTGTATACCTAAACGACCACTCAGCCCTCCAACTCACCCTTGCCTGC 240  
Qy 240 CAGACCTCACCACATCTTGTGAGCTCAAACTCAGCCGACCTAACTCAACCAATCCC 299  
Db 241 CAGACCTCACCACATCTTGTGAGCTCAAACTCAGCCGACCTAACTCAACCAATCCC 300  
Qy 300 AAGCTAAACTAATCTGAAACTTTTAAAGTAAACCCAGTCTTAAACCTTAACTAGCCCA 359  
Db 301 AAGCTAAACTAATCTGAAACTTTTAAAGTAAACCCAGTCTTAAACCTTAACTAGCCCA 360  
Qy 360 TGCAATATATCTACCCCTAGCCAAACCTTAACTGCTTTCCTGCTCCAGTCCCAAGTGTCCACT 419  
Db 361 TGCAATATATCTACCCCTAGCCAAACCTTAACTGCTTTCCTGCTCCAGTCCCAAGTGTCCACT 420

QY 420 GAATCCTGACCTTGGTCTCTCACTGAAATCCAGAAAGACATATTTCCCACTGCCCCA 479  
Db 421 GAATCCTGACCTTGGTCTCTCACTGAAATCCAGAAAGACATATTTCCCACTGCCCCA 480  
QY 480 TCCCTCTTTACAGCACCAACCTGCTCTGACCTCTGCTGATCTGCTGGATGCTCCAAAC 539  
Db 481 TCCCTCTTTACAGCACCAACCTGCTCTGACCTCTGCTGATCTGCTGGATGCTCCAAAC 540  
QY 540 TCTGAGTGCCTATGAGCAACCAAGCCGACTCGTCAAAATGCACCTCTCTCCCTTCTCTGTC 599  
Db 541 TCTGAGTGCCTATGAGCAACCAAGCCGACTCGTCAAAATGCACCTCTCTCCCTTCTCTGTC 600  
QY 600 CCCACCTTGCAGGCTGATGGAAGCCCTCATTAAGTCAAACTTTCCCACTCAACAC 659  
Db 601 CCCACCTTGCAGGCTGATGGAAGCCCTCATTAAGTCAAACTTTCCCACTCAACAC 660  
QY 660 CAAGAACGGGCTGAACCTCCACACTGCCACCGTTCCCTGAGAGTGAGCACTAAATCTCCT 719  
Db 661 CAAGAACGGGCTGAACCTCCACACTGCCACCGTTCCCTGAGAGTGAGCACTAAATCTCCT 720  
QY 720 TCAATCTAACCCACCTTACATTTCCCACTCACTGAGGAATCAATCTAGAAATATACCCAA 779  
Db 721 TCAATCTAACCCACCTTACATTTCCCACTCACTGAGGAATCAATCTAGAAATATACCCAA 780  
QY 780 AACTAAGCCCCATAGGAGCCGACCTCTAGTGTCTTACCCCTATACCTTCTCTATG 839  
Db 781 AACTAAGCCCCATAGGAGCCGACCTCTAGTGTCTTACCCCTATACCTTCTCTATG 840  
QY 840 GGTGAGTCTGTCTTTGGGGCCGCTCTCTCTGCTCTCTCTCTTCTCTCTCTCTCTCTCT 899  
Db 841 GGTGAGTCTGTCTTTGGGGCCGCTCTCTCTGCTCTCTCTCTTCTCTCTCTCTCTCTCT 900  
QY 900 CAGCTGCGAGCTGTGACATGTGTCTTCCACCTCTGACTCCCTCAAGCTGCAGTG 959  
Db 901 CAGCTGCGAGCTGTGACATGTGTCTTCCACCTCTGACTCCCTCAAGCTGCAGTG 960  
QY 960 GGACTGGAAGACTGCGAGGAGCTAGGCTAGCACTGGAACACAGGAGCTGACCTGACG 1019  
Db 961 GGACTGGAAGACTGCGAGGAGCTAGGCTAGCACTGGAACACAGGAGCTGACCTGACG 1020  
QY 1020 TCCCTAGGCTTGGCCCCCTCTCTCTCATGTACACATATATATGTTGGCACACACAGT 1079  
Db 1021 TCCCTAGGCTTGGCCCCCTCTCTCTCATGTACACATATATATGTTGGCACACACAGT 1080  
QY 1080 GGCAACATGCAAGACTCTCTGAGTGAACACAGATTCATTTCTCAAGTATCTACTGA 1139  
Db 1081 GGCAACATGCAAGACTCTCTGAGCTGACACAGATTCATTTCTCAAGTATCTACTGA 1140  
QY 1140 TAGACACTCATGCGTGCCAAAGTCTCTCATCTCAACATACACATGCTCTCTCTCTCTCC 1199  
Db 1141 TAGACACTCATGCGTGCCAAAGTCTCTCATCTCAACATACACATGCTCTCTCTCTCTCC 1200  
QY 1200 CGTCTTGGCAGAGTGTTCCTCTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 1259  
Db 1201 CGTCTTGGCAGAGTGTTCCTCTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 1260  
QY 1260 CTCACCCCAACAGCCAGGCTGGGACAGACACCTGAGGGGCTGACGCTGCTTCCC 1319  
Db 1261 CTCACCCCAACAGCCAGGCTGGGACAGACACCTGAGGGGCTGACGCTGCTTCCC 1320  
QY 1320 CGTGTGGGCGCGGCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 1379  
Db 1321 CGTGTGGGCGCGGCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 1380  
QY 1380 GGATGTAGCTGTGTCAGCTGTCTGGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAATC 1439  
Db 1381 GGATGTAGCTGTGTCAGCTGTCTGGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAATC 1440  
QY 1440 GCAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTGACCCGCTACC 1499  
Db 1441 GCAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTGACCCGCTACC 1500  
QY 1500 TGGAGCAACCACTCCGAGCTTGGCTGGGACCTATGTAGTATCCAGCGTAGGAATCTGG 1559

Db 1501 TGAGCACCAACTCCGACCTTGGCTGGGACCTATGTAGATATCCAGCGTAGGAATCTGG 1560  
QY 1560 GAGTTGGGAGAGTAGAGAGTTGGGAAAGACAGTCTTAAACGCTGGAGGGTTCTGGTAA 1619  
Db 1561 GAGTTGGGAGAGTAGAGAGTTGGGAAAGACAGTCTTAAACGCTGGAGGGTTCTGGTAA 1620  
QY 1620 ATGATGGGAGTAGAGGGGCTCTTTGGCTCCCAACAGTCCCTGCTCTCTCTCTCTCTCT 1679  
Db 1621 ATGATGGGAGTAGAGGGGCTCTTTGGCTCCCAACAGTCCCTGCTCTCTCTCTCTCTCT 1680  
QY 1680 GGCCTTCCCTCTTAGTGGCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1739  
Db 1681 GGCCTTCCCTCTTAGTGGCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1740  
QY 1740 GGGCAGGCTGCAACCCGCTTGGCCATTTGCCCATCTGGCTGCCAGCCAGCCGCGCGC 1799  
Db 1741 GGGCAGGCTGCAACCCGCTTGGCCATTTGCCCATCTGGCTGCCAGCCAGCCGCGCGC 1800  
QY 1800 CTCCCTCTGGGGCGGGGAGTCTCTCTGTTTACACCGTGTGTGTGTGTCTCTCTCTCTCT 1859  
Db 1801 CTCCCTCTGGGGCGGGGAGTCTCTCTGTTTACACCGTGTGTGTGTGTCTCTCTCTCTCT 1860  
QY 1860 GGGCGGGTGTGGTGGGACAGAGGGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1919  
Db 1861 GGGCGGGTGTGGTGGGACAGAGGGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1920  
QY 1920 CTGCCCCCAGACTGGGGCCCTCTCTCTGACCCAGGGGCTCTCTCTCTCTCTCTCTCTCT 1979  
Db 1921 CTGCCCCCAGACTGGGGCCCTCTCTCTGACCCAGGGGCTCTCTCTCTCTCTCTCTCTCT 1980  
QY 1980 TCCCATCTAGTGGGCTCTCTAGGGGGCTCATGGGGGAGGGGACTGTAGGGAAACCCAG 2039  
Db 1981 TCCCATCTAGTGGGCTCTCTAGGGGGCTCATGGGGGAGGGGACTGTAGGGAAACCCAG 2040  
QY 2040 GCAGTAGTGGCAGGGGTTTAGGGTGTGGATGGAGTTATGCTGTAAAGATTTGGGGTG 2099  
Db 2041 GCAGTAGTGGCAGGGGTTTAGGGTGTGGATGGAGTTATGCTGTAAAGATTTGGGGTG 2100  
QY 2100 GTCCAGAGTGTTCAGAGAGCCAGGAGAGAGGAGGAGGGTTCGAGAGCCGAGGAC 2159  
Db 2101 GTCCAGAGTGTTCAGAGAGCCAGGAGAGAGGAGGAGGGTTCGAGAGCCGAGGAC 2160  
QY 2160 CATGGGAAACCGGCT 2219  
Db 2161 CATGGGAAACCGGCT 2220  
QY 2220 GCCAGGAAAGAAAGGAGAGAGTGGCGGGGAGCTGGCTCCAGCCCCAGGATACACC 2279  
Db 2221 GCCAGGAAAGAAAGGAGAGAGTGGCGGGGAGCTGGCTCCAGCCCCAGGATACACC 2280  
QY 2280 GAGGAAATAGTGTCT 2339  
Db 2281 GAGGAAATAGTGTCT 2340  
QY 2340 ATCCAGGCT 2399  
Db 2341 ATCCAGGCT 2400  
QY 2400 TTCCCTGGGCCCCAGCGCTCCCGAGGGTGGAAAGGGCTCTGCGCTCTCTCTCTCTCTATAC 2459  
Db 2401 TTCCCTGGGCCCCAGCGCTCCCGAGGGTGGAAAGGGCTCTGCGCTCTCTCTCTCTCTATAC 2460  
QY 2460 CATGCTGTCTTCCATAGCT 2519  
Db 2461 CATGCTGTCTTCCATAGCT 2520  
QY 2520 TCTGCAACCTGCT 2579  
Db 2521 TCTGCAACCTGCT 2580  
QY 2580 CTCCCAAGCCCTCAGCTGTGGGCTCTGCTGTCTGAGCGCAATGGGGCTCTGCTTCTCTCTCT 2639



QY 4800 ATTAAACAATAAATGCTTGGCCCAAGAACTTAGTCTCTTTGGCCAGCCTCAACCCCTCCTG 4859  
 Db ATTTACAAATAAATGCTTGGCCCAAGAACTTAGTCTCTTTGGCCAGCCTCAACCCCTCCTG 4860  
 QY 4860 GTGCTCATCAGACTCTTGCACCCCTGGCTCCCACTCCCTGCTTGGCTGCTGGAGCTG 4919  
 Db GTGCTCATCAGACTCTTGCACCCCTGGCTCCCACTCCCTGCTTGGCTGCTGGAGCTG 4920  
 QY 4920 CACAGAGCTCTGGGAAGAGGCGCTCTTCTCCCTCCCGCACTGGGGCGGATGGGGCACCCTCAG 4979  
 Db CACAGAGCTCTGGGAAGAGGCGCTCTTCTCCCTCCCGCACTGGGGCGGATGGGGCACCCTCAG 4980  
 QY 4980 ACTTACCACCTGCTGCTGCGACCAACACCCCTTGATCCCTCAGTCCTCCACACAGCTT 5039  
 Db ACTTACCACCTGCTGCTGCGACCAACACCCCTTGATCCCTCAGTCCTCCACACAGCTT 5040  
 QY 5040 CTGTCCACCCAGGTTTCCCTCACCCACCTTTGCTAAGTCTTCTCTCA 5087  
 Db CTGTCCACCCAGGTTTCCCTCACCCACCTTTGCTAAGTCTTCTCTCA 5088

## RESULT 5

AAA88546

ID AAA88546 standard; DNA; 1790 BP.

XX AAA88546;

AC AAA88546;

XX 22-JAN-2001 (first entry)

XX Human interleukin-B60 (IL-B60) gene.

XX Interleukin-B60; IL-B60; human; cytokine; chromosome 11;

KW cytokine-like factor-1; haematopoietic; inflammation;

KW antiinflammatory; autoimmune disease; therapy; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FH CDS 162..809

FT sig\_peptide 162..212

FT mat\_peptide 213..806

FT /\*tag= c

XX WO200053631-A1.

XX 14-SEP-2000.

XX 09-MAR-1999; 2000WO-US06182.

XX 11-MAR-1999; 99US-0267901.

XX (SCHE ) SCHERING CORP.

XX Oppmann B, Timans JC, Kastelein RA, Bazan JF;

XX WPI; 2000-587426/55.

XX P-PSDB; AAB19686.

XX Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes,

XX polypeptides, and nucleic acids, useful in research, diagnosis and for

XX treating inflammatory and autoimmune disorders -

XX Claim 17; Page 15-16; 97pp; English.

XX The present sequence is that of DNA encoding human interleukin-B60

XX (IL-B60, see AAB19586), a novel, small soluble cytokine-like protein

XX of 198 amino acids that exhibits structural motifs characteristic

XX of a member of the long-chain cytokines, and which shows homology

XX to granulocyte colony stimulating factor and interleukin-6. IL-B60

XX may have either stimulatory or inhibitory effects on haematopoietic

XX cells, including e.g. lymphoid cells, such as T-cells, B-cells,

CC natural killer cells, macrophages, dendritic cells, haematopoietic  
 CC progenitors, etc. Methods are provided for modulating the  
 CC physiology or development of a cell or tissue culture cells by  
 CC contacting the cell with an agonist or antagonist of IL-B60 or an  
 CC agonist of antagonist of a complex of mature IL-B60 and its  
 CC partner, cytokine-like factor-1 (CLF-1, see AAB19588). The  
 CC IL-B60/CLF-1 cytokine serves as a key physiological factor in motor  
 CC neuron development and regeneration. IL-60B, its agonists and  
 CC antagonists may be used to treat inflammatory or autoimmune  
 CC disorders and also for drug screening. The IL60B gene maps to  
 CC human chromosome 11.  
 XX

SQ Sequence 1790 BP; 381 A; 560 C; 474 G; 375 T; 0 other;

Query Match 28.8%; Score 1464.4; DB 21; Length 1790;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCCTGGGG 3422

Db ATCTGACTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCCTGGGG 372

QY 3423 CAGAGACTCTGCCAGGCCCACTGTGTGACTTGGAGGTGTGGCAAGCTCAATGACAAC 3482

Db CAGAGACTCTGCCAGGCCCACTGTGTGACTTGGAGGTGTGGCAAGCTCAATGACAAC 432

QY 3483 TGGGCTGACCCAGAACTACGAGGCTTACGAGCCCTTCTGTGTACTTGGTGGCTCA 3542

Db TGGGCTGACCCAGAACTACGAGGCTTACGAGCCCTTCTGTGTACTTGGTGGCTCA 492

QY 3543 ACCGTGAGGCTGCCCACTGTGTGAGCTGGCGCCAGCTTGGCCCACTTCTGCAACCACTCC 3602

Db ACCGTGAGGCTGCCCACTGTGTGAGCTGGCGCCAGCTTGGCCCACTTCTGCAACCACTCC 552

QY 3603 AGGCTCTGTGGGAGCAATTCGGGGCGTATGGAGCTCTGGGTATCCCATGCCCCAGC 3662

Db AGGCTCTGTGGGAGCAATTCGGGGCGTATGGAGCTCTGGGTATCCCATGCCCCAGC 612

QY 3663 CGCTGCTGGGAGCAATTCGGGGCGTATGGAGCTCTGGGTATCCCATGCCCCAGC 3722

Db CGCTGCTGGGAGCAATTCGGGGCGTATGGAGCTCTGGGTATCCCATGCCCCAGC 672

QY 3723 AGATGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGTGGCTGTGGCTGTGGCTGTGG 3782

Db AGATGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGTGGCTGTGGCTGTGGCTGTGG 732

QY 3783 ACTTCAACCCGGCTCAAGAAGAGATGAGCCTCCAGAGCTGCAGTCACTGTGACCTGTG 3842

Db ACTTCAACCCGGCTCAAGAAGAGATGAGCCTCCAGAGCTGCAGTCACTGTGACCTGTG 792

QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTGCTCCCTTCCCTTCCCTTCCCTTCCCTTCC 3902

Db GGGCTCATGGCTTCTGACTTCTGACCTTCTGCTCCCTTCCCTTCCCTTCCCTTCCCTTCC 852

QY 3903 CACTTTGTGAGAGCCAGCCTGTATGCCAACACCTGTGTGAGCCAGGAGACAGAGCTGTG 3962

Db CACTTTGTGAGAGCCAGCCTGTATGCCAACACCTGTGTGAGCCAGGAGACAGAGCTGTG 912

QY 3963 AGCTCTGGGCTTCTGACCTGTGGACCGGCTGGGCTGTGTATGCGATCAGGCTGTCTCTCC 4022

Db AGCTCTGGGCTTCTGACCTGTGGACCGGCTGGGCTGTGTATGCGATCAGGCTGTCTCTCC 972

QY 4023 CACCTCCCAAGGCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTGTCTCTGT 4082

Db CACCTCCCAAGGCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTGTCTCTGT 1032

QY 4083 TCTACAGGAAGTCACTGTCCAGGAGGTGTGAAGTGTGTGAGTTGGTGTGACAGGCGCTCA 4142

Db TCTACAGGAAGTCACTGTCCAGGAGGTGTGAAGTGTGTGAGTTGGTGTGACAGGCGCTCA 1092

QY 4143 TGGCTCTCTGCTTCTTCCCTACCACTTGGGAGGTGCCACCCAGCCCTCTCAGGTGGCACA 4202

Db TGGCTCTCTGCTTCTTCCCTACCACTTGGGAGGTGCCACCCAGCCCTCTCAGGTGGCACA 1152

QY 4203 TCTGAGGCGAGGGTTGAGGGGCCACACACATGCTTTCTGGGTGAAGCCCTTT 4262  
Db 1153 TCTGAGGCGAGGGTTGAGGGGCCACACACATGCTTTCTGGGTGAAGCCCTTT 1212  
QY 4263 GGCTCCGCCACTCTCTCTGGATGGGTGTGCTCCCTTATCCCAATCACTCTATACATC 4322  
Db 1213 GGCTCCGCCACTCTCTCTGGATGGGTGTGCTCCCTTATCCCAATCACTCTATACATC 1272  
QY 4323 CAATTGAGGAACAACATGTTGGGCAATTTCTACAAAAAGAGATGAGATTAACAGTGCA 4382  
Db 1273 CAATTGAGGAACAACATGTTGGCAATTTCTACAAAAAGAGATGAGATTAACAGTGCA 1332  
QY 4383 GGGTGGGGTCTGCAATGGAGTGGCTTAAACACAGAGAGAAATCTGAAAGCACAG 4442  
Db 1333 GGGTGGGGTCTGCAATGGAGTGGCTTAAACACAGAGAGAAATCTGAAAGCACAG 1392  
QY 4443 GGGCAGGACAGACACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAAAC 4502  
Db 1393 GGGCAGGACAGACACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAAAC 1452  
QY 4503 CCGAGCTGAGCATCAGGACCTTGCCTCGAATTTCTCCAGTATTAGGTGCTCTCTC 4562  
Db 1453 CCGAGCTGAGCATCAGGACCTTGCCTCGAATTTCTCCAGTATTAGGTGCTCTCTC 1512  
QY 4563 TGCCCCCTTTCCAGGGTATCTGTGGTTCGCCAGGCTGGGGAGGCAACCATAGCCACAC 4622  
Db 1513 TGCCCCCTTTCCAGGGTATCTGTGGTTCGCCAGGCTGGGGAGGCAACCATAGCCACAC 1572  
QY 4623 CACAGGATTTCTGAAAGTTTACAATGCAATGAGTATTTGGGGTGTAGGGTGGCAGTCC 4682  
Db 1573 CACAGGATTTCTGAAAGTTTACAATGCAATGAGTATTTGGGGTGTAGGGTGGCAGTCC 1632  
QY 4683 CCAAGCCCTGCCCCCAGCCACCCACCATCATGACTCTAAGTGTGTGTTAATATTT 4742  
Db 1633 CCAAGCCCTGCCCCCAGCCACCCACCATCATGACTCTAAGTGTGTGTTAATATTT 1692  
QY 4743 ATTTATTTGGAGATGTTATTTATTTAGATATTTATTTAGAGATTTCTATTCTTGATT 4802  
Db 1693 ATTTATTTGGAGATGTTATTTATTTAGATATTTATTTAGAGATTTCTATTCTTGATT 1752  
QY 4803 AACAAATAAAATGCTTGGCCCCAGAAC 4828  
Db 1753 AACAAATAAAATGCTTGGCCCCAGAAC 1778

## RESULT 6

AXX16161  
ID AAX16161 standard; DNA; 1710 BP.

XX AAX16161;

XX AAX16161;

DT 28-APR-1999 (first entry)

DE Human cardiotrophin-like cytokine encoding DNA.

XX Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;  
KW CLC; IL-6; diagnosis; detection; immune system-related disorder;  
KW cancer; cardiac disorder; heart failure; hypertension; cancer;  
KW autoimmune disorder; infection; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
FH CDS 46..723

FT sig\_peptide /\*tag= a

FT /\*tag= b

FT mat\_peptide 130..720

FT /\*tag= c  
/product= "cardiotrophin-like cytokine"

XX PN WO99004157A1.

XX 07-JAN-1999.  
PD 29-JUN-1998; 98WO-US13129.  
XX 30-JUN-1997; 97US-0051311.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Ruben SM, Shi Y;  
PI WPI; 1999-095678/38.  
XX P-PSDB; AAW94466.  
DR New isolated cardiotrophin-like cytokine nucleic acid - used to  
DR develop products for treating cardiac and immune system disorders,  
DR e.g. heart failure, hypertension, cancers, autoimmune disorders and  
XX PT infections  
XX Claim 2; Fig 1; 103pp; English.  
XX The present invention relates to a novel cardiotrophin-like cytokine  
CC (CLC) protein which is a member of the interleukin 6 (IL-6) cytokine  
CC family. The present sequence encodes the human CLC protein. The  
CC present invention also describes screening methods for identifying  
CC agonists and antagonists of CLC activity, as well as methods for  
CC detecting cardiac and immune system-related disorders and  
CC therapeutic methods for treating cardiac and immune system-related  
CC disorders, e.g. heart failure, hypertension, cancers, autoimmune  
CC disorders and infections.  
XX Sequence 1710 BP; 370 A; 530 C; 448 G; 362 T; 0 other;

Query Match 28.8%; Score 1463.4; DB 20; Length 1710;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1464; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3363 AGCTGAACTACTCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGGG 3422  
Db 227 ATCTGAACTACTCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGGG 286  
QY 3423 CAGAGACTCTGCCAGGGCCACTGTGTGACTTTGGAGGTGTGGGAGGCTCAATGACAAAC 3482  
Db 287 CAGAGACTCTGCCAGGGCCACTGTGTGACTTTGGAGGTGTGGGAGGCTCAATGACAAAC 346  
QY 3483 TGCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTTACTTGTGGCTCA 3542  
Db 347 TGCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTTACTTGTGGCTCA 406  
QY 3543 ACCGTGAGGCTGCCACTGCTGAGCTGCGCGGAGCTTGGCCCACTTCTGACAGCTCC 3602  
Db 407 ACCGTGAGGCTGCCACTGCTGAGCTGCGCGGAGCTTGGCCCACTTCTGACAGCTCC 466  
QY 3603 AGGCGCTGCTGGGAGCAATTCGGGCGTTCATGGAGCTCTGGGCTACCCACTGCCCCAGC 3662  
Db 467 AGGCGCTGCTGGGAGCAATTCGGGCGTTCATGGAGCTCTGGGCTACCCACTGCCCCAGC 526  
QY 3663 CGCTGCTGGGAGCTGAACCCCACTTGGACTCTCTGGGCTGCCCCACAGTACTTCTCCAGA 3722  
Db 527 CGCTGCTGGGAGCTGAACCCCACTTGGACTCTCTGGGCTGCCCCACAGTACTTCTCCAGA 586  
QY 3723 AGATGGAGCACTTCTGGCTGCTGAAGGAGCTGCAGACCTTGGCTGTGGCGCTGGGCAAGG 3782  
Db 587 AGATGGAGCACTTCTGGCTGCTGAAGGAGCTGCAGACCTTGGCTGTGGCGCTGGGCAAGG 646  
QY 3783 ACTTCAACCGGCTCAAGAAAGATGAGGCTTCCAGAGCTCAGTCACTGACCTGG 3842  
Db 647 ACTTCAACCGGCTCAAGAAAGATGAGGCTTCCAGAGCTCAGTCACTGACCTGG 706  
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGCTCTTCCCTTCAACCTTCAACCTGCTCC 3902  
Db 707 GGGCTCATGGCTTCTGACTTCTGACTTCTGCTCTTCCCTTCAACCTTCAACCTGCTCC 766

QY 3903 CACTTTGTGAGAGCCGCTGTATGCCAACACCTGTTGAGCCAGGAGACAGAGCTGTG 3962  
Db |||||  
QY 767 CACTTTGTGAGAGCCGCTGTATGCCAACACCTGTTGAGCCAGGAGACAGAGCTGTG 826  
Db |||||  
QY 3963 AGCCTCTGGCCCTTCTCCCTGGACCGGCTGGGGTGGATGCGATCAGCCCTGCTCTCTCC 4022  
Db |||||  
QY 827 AGCCTCTGGCCCTTCTCCCTGGACCGGCTGGGGTGGATGCGATCAGCCCTGCTCTCTCC 886  
Db |||||  
QY 4023 CACTTCCCAAGGCTACCGAGCTGGGGAGGAGGTACAGTAGGCGCTCTCTCTCTCTCT 4082  
Db |||||  
QY 887 CACTTCCCAAGGCTACCGAGCTGGGGAGGAGGTACAGTAGGCGCTCTCTCTCTCTCT 946  
Db |||||  
QY 4083 TCTACAGGAAGTCACTGCTCGAGGAGTGTGAAGTGGTTGAGTGGTGGTGGTGGTGGT 4142  
Db |||||  
QY 947 TCTACAGGAAGTCACTGCTCGAGGAGTGTGAAGTGGTTGAGTGGTGGTGGTGGTGGT 1006  
Db |||||  
QY 4143 TGGCT 4202  
Db |||||  
QY 1007 TGGCT 1066  
Db |||||  
QY 4203 TCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4262  
Db |||||  
QY 1067 TCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1126  
Db |||||  
QY 4263 GGCCTGCCCCACTCTCTCTGAGTGGTGTCTCCCTTATCCCAATCACTCTATATATC 4322  
Db |||||  
QY 1127 GGCCTGCCCCACTCTCTCTGAGTGGTGTCTCCCTTATCCCAATCACTCTATATATC 1186  
Db |||||  
QY 4323 CAATTACAGGAACCAACATGCTGGCAATCTTACACAAAGAGATGAGATTAACAGTGCA 4382  
Db |||||  
QY 1187 CAATTACAGGAACCAACATGCTGGCAATCTTACACAAAGAGATGAGATTAACAGTGCA 1246  
Db |||||  
QY 4383 GGGTTGGGTTCTGATTGGAGTGCCTTATAAACCAGAGAGAAATATCTGAAAGCACAG 4442  
Db |||||  
QY 1247 GGGTTGGGTTCTGATTGGAGTGCCTTATAAACCAGAGAGAAATATCTGAAAGCACAG 1306  
Db |||||  
QY 4443 GGGCAGGAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 4502  
Db |||||  
QY 1307 GGGCAGGAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 1366  
Db |||||  
QY 4503 CCGAGCTGAGCATCAGAGACCTTGCTCGAATTTCTTCCAGTATTACGGTCCCTCTCTC 4562  
Db |||||  
QY 1367 CCGAGCTGAGCATCAGAGACCTTGCTCGAATTTCTTCCAGTATTACGGTCCCTCTCTC 1426  
Db |||||  
QY 4563 TGCCCCCTTTCCAGGATCTGTTGGTTCAGGCTGGGAGGGGCAACCATAGCCACAC 4622  
Db |||||  
QY 1427 TGCCCCCTTTCCAGGATCTGTTGGTTCAGGCTGGGAGGGCAACCATAGCCACAC 1486  
Db |||||  
QY 4623 CACAGGATTTCTGAAAGTTTACATGCAAGTACATTTTGGGTTGAGGTTGAGGTTGAGG 4682  
Db |||||  
QY 1487 CACAGGATTTCTGAAAGTTTACATGCAAGTACATTTTGGGTTGAGGTTGAGGTTGAGG 1546  
Db |||||  
QY 4683 CCAAGGCTTGGCCCCCAGCCCAACCACTCATGCTCTAAGTGTGTTGATTAATATTT 4742  
Db |||||  
QY 1547 CCAAGGCTTGGCCCCCAGCCCAACCACTCATGCTCTAAGTGTGTTGATTAATATTT 1606  
Db |||||  
QY 4743 ATTTATTTGAGATGTTATTTATTTAGATGATTTATTTGAGATTTCTATTTCTGATTT 4802  
Db |||||  
QY 1607 ATTTATTTGAGATGTTATTTATTTAGATGATTTATTTGAGATTTCTATTTCTGATTT 1666  
Db |||||  
QY 4803 AACAAATAAATGCTTGGCCCCAGAA 4827  
Db |||||  
QY 1667 AACAAATAAATGCTTGGCCCCAGAA 1691  
Db |||||

## RESULT 7

AAK51548

ID AAK51548 standard; cDNA; 1008 BP.

XX

AC AAK51548;

XX

DT 06-NOV-2001 (first entry)

XX

Human polynucleotide SEQ ID NO 93.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
tissue growth factor; immunomodulatory; cancer; leukaemia;  
nervous system disorder; arthritis; inflammation; ss.

Homo sapiens.

W0200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0398075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Auandi V, Zhou P, Xu C, Cao Y, Ma Y;  
XX Zhao QA, Wang D, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
XX WPI; 2001-476283/51.

DR P-PSDB; AAM78415.

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
XX useful in diagnosis and gene therapy -  
XX Claim 1; Page 711-712; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.

XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
XX (AAM80020) are omitted as the relevant pages from the sequence listing  
XX were missing at the time of publication.

XX SQ Sequence 1008 BP; 183 A; 374 C; 247 G; 204 T; 0 other;  
Query Match 10.5%; Score 535.4; DB 22; Length 1008;  
Best Local Similarity 99.8%; Pred. No. 5.9e-123;  
Matches 536; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGGG 3422  
Db |||||

QY 472 ATCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGGG 531  
Db |||||

QY 3423 CAGAGACTCTGCCAGGGCCACTGTTGACCTTGGAGGTGTGGCGAAGCTCAATGACAAAC 3482  
Db |||||

QY 532 CAGAGACTCTGCCAGGGCCACTGTTGACCTTGGAGGTGTGGCGAAGCTCAATGACAAAC 591  
Db |||||

QY 3483 TGGGGCTGAGCCAGAACTACGAGGCTTACAGCCACTTCTGTGTTACTTGGTGGCTTCA 3542  
Db |||||

QY 592 TGGGGCTGAGCCAGAACTACGAGGCTTACAGCCACTTCTGTGTTACTTGGTGGCTTCA 651  
Db |||||

QY 3543 ACCGTGAGGTGCACTGTGAGCTGCGCGCAGCTTCTGGCCACTTCTGACACACGCTCC 3602  
Db |||||



```
Db 652 ACCGTCAGGCTGCCACTGCTGAGCTGCGCGCAGCGCTGGCCACACTTCTGCAACAGCCTCC 711
Qy 3603 AGGGCTCTGGGACGACTTGGGGCGTCAATGGCAGCTCTGGGCTACCCACTGCCCCAGC 3662
Db 712 AGGGCTCTGGGACGACTTGGGGCGTCAATGGCAGCTCTGGGCTACCCACTGCCCCAGC 771
Qy 3663 CGCTGCTGGGACTGAACCCCACTTGGACTCTGGCCCTGCCACAGTGAATCTCTCCAGA 3722
Db 772 CGCTGCTGGGACTGAACCCCACTTGGACTCTGGCCCTGCCACAGTGAATCTCTCCAGA 831
Qy 3723 AGATGGACGACTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGGCGCTGGCCCAAG 3782
Db 832 AGATGGACGACTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGGCGCTGGCCCAAG 891
Qy 3783 ACTTCAACCGGCTCAAGAAGAGATGACGCTCCAGCAGCTGCAGTCAACCTGCACCTGG 3842
Db 892 ACTTCAACCGGCTCAAGAAGAGATGACGCTCCAGCAGCTGCAGTCAACCTGCACCTGG 951
Qy 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3899
Db 952 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1008

RESULT 9
AAD04201
ID AAD04201 standard; cDNA; 729 BP.
XX AC AAD04201;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human cardiotrophin-like cytokine (CLC) cDNA.
XX
KW Human; biologically active complex; haemopoietin receptor; NR6;
KW cardiotrophin-like cytokine; CLC; therapy; prophylaxis; proliferation;
KW differentiation; cell survival; neurotrophic activity; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 11..679
FT /*tag= a
FT /product= "Human cardiotrophin-like cytokine (CLC)
FT protein"
FT /note= "CDS does not include stop codon"
FT /partial
FT sig_peptide 11..91
FT /*tag= b
FT mat_peptide 92..679
FT /*tag= c
FT /product= "Human mature cardiotrophin-like cytokine
FT (CLC) protein"
XX
FN WO200127157-A1.
XX
PD 19-APR-2001.
XX
PF 06-OCT-2000; 2000WO-AU01216.
XX
PR 08-OCT-1999; 99AU-0003327.
PR 12-MAY-2000; 2000AU-0007489.
XX
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
XX Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;
XX Nakata Y, Hasegawa M;
XX
XX WPI; 2001-281978/29.
XX
XX P-PSDB; AAE00828.
XX
XX New biologically active complex comprising NR6 and
XX cardiotrophin-like-cytokine, for facilitating proliferation,
XX differentiation and/or survival of a cell -
```

```
XX PS
XX Claim 31; Page 112-114; 123pp; English.
XX
CC The present invention relates to a biologically active complex comprising
CC a haemopoietin receptor, NR6 and cardiotrophin-like cytokine (CLC).
CC The complex is useful in the manufacture of a medicament for the
CC treatment and/or prophylaxis of a subject, as it is involved in
CC facilitating proliferation, differentiation and/or survival of a cell.
CC The complex or its components have neurotrophic activity. The present
CC sequence is human cardiotrophin-like cytokine (CLC) cDNA.
XX
XX Sequence 729 BP; 132 A; 261 C; 196 G; 140 T; 0 other;
SQ
Query Match 10.4%; Score 529.4; DB 22; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.6e-121;
Matches 530; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3363 AGTGTAACTACTGGGGCCCGCTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGGG 3422
Db 192 ATCTGAACACTACCTGGGGCCCGCTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGGG 251
Qy 3423 CAGAGACTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGCAGAGCCTCAATGACAAAC 3482
Db 252 CAGAGACTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGCAGAGCCTCAATGACAAAC 311
Qy 3483 TGGCGCTGACCCAGAACTACGAGGCCCTACAGCCACCTCTGTGTACTTGGGTGGCCCTCA 3542
Db 312 TGGCGCTGACCCAGAACTACGAGGCCCTACAGCCACCTCTGTGTACTTGGGTGGCCCTCA 371
Qy 3543 ACCGTACGGCTGCGCACTGCTGAGCTGGCGCGCAGCCTGGCCACTTCTGCAACCCAGCCTCC 3602
Db 372 ACCGTACGGCTGCGCACTGCTGAGCTGGCGCGCAGCCTGGCCACTTCTGCAACCCAGCCTCC 431
Qy 3603 AGGGCTCTGGGACGACTTGGCGGCTCATGGCAGCTCTGGGCTACCCACTGCCCCAGC 3662
Db 432 AGGGCTCTGGGACGACTTGGCGGCTCATGGCAGCTCTGGGCTACCCACTGCCCCAGC 491
Qy 3663 CGCTGCTGGGACTGAACCCCACTTGGACTCTCTGGCCCTGCCACAGTGAATCTCTCCAGA 3722
Db 492 CGCTGCTGGGACTGAACCCCACTTGGACTCTCTGGCCCTGCCACAGTGAATCTCTCCAGA 551
Qy 3723 AGATGGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGGCGCTGGCCCAAG 3782
Db 552 AGATGGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGGCGCTGGCCCAAG 611
Qy 3783 ACTTCAACCGGCTCAAGAAGAGATGACGCTCCAGCAGCTGCAGTCAACCTGCACCTGG 3842
Db 612 ACTTCAACCGGCTCAAGAAGAGATGACGCTCCAGCAGCTGCAGTCAACCTGCACCTGG 671
Qy 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3893
Db 672 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 722

RESULT 9
AAH74484
ID AAH74484 standard; DNA; 881 BP.
XX AC AAH74484;
XX
DT 15-OCT-2001 (first entry)
XX
DE Nucleotide sequence of a human NNT-1 protein.
XX
KW NNT-1; CLF-1; scNTPRalpha; nervous system; neuron; nervous system;
KW neuro-muscular function; tumour; immune system; haematopoietic system;
KW reproductive system; liver; skeletal muscle; neurodegenerative disease;
KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;
KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;
KW blastocyst implantation; thrombosis; retinal disease;
KW retinal pigmentosa; ss.
XX
XX Homo sapiens.
OS
```



XX Key Location/Qualifiers  
FH CDS 174..851  
FT /\*tag= a  
FT /\*product= "NNT-1"  
XX WO200155172-A2.  
XX PD 02-AUG-2001.  
XX 26-JAN-2001; 2001WO-FR00253.  
XX PF 27-JAN-2000; 2000FR-0001035.  
XX PR 12-OCT-2000; 2000FR-0013089.  
XX (FABR ) FABRE MEDICAMENT SA PIERRE.  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;  
XX WPI; 2001-488773/53.  
DR P-PSDB; AAG63543.  
XX  
XX A complex comprising a NNT-1 protein and a CLF-1 and/or SCNTFRalpha  
PT protein useful to treat neurodegenerative disease including Parkinson's  
PT and Huntington's, obesity and cancer -  
XX Disclosure; Page 57-58; 67pp; French.  
XX  
XX The present sequence encodes a human NNT-1 protein. The specification  
CC describes a complex comprising a NNT-1 protein and a CLF-1 and/or  
CC SCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate  
CC activity of the SCNTFRalpha/gp130/LiPRbeta receptor complex, or to  
CC induce phosphorylation of the tyrosine of gp130 and LiPRbeta,  
CC particularly where cells expressing the receptor complex are in the  
CC central or peripheral nervous system, in neurons implicated in  
CC neuro-muscular function or in skeletal muscle. The complex or  
CC antibodies are also used to decrease the survival, growth or  
CC proliferation of tumour cells or to facilitate the proliferation and/or  
CC inhibit differentiation of cells stocks. The complex is also used to  
CC modulate activity of the gp130/LiPRbeta receptor or cells expressing  
CC that receptor, particularly those cells implicated in the immune,  
CC haematopoietic, nervous or reproductive system, the liver or skeletal  
CC muscle. Molecules of the invention may be used to prevent or treat  
CC neurodegenerative diseases including amyotrophic lateral sclerosis,  
CC Parkinson's and Huntington's disease, to repair or regenerate nervous  
CC or muscular tissue or to maintain muscular mass in paralysis patients.  
CC They may also be used to treat cancer, obesity and associated diseases,  
CC and to improve fertility, particularly to avoid endometriosis and/or  
CC assist blastocyst implantation, thrombosis, or retinal disease,  
CC particular retinal pigmentosis.  
XX  
XX Query Match 10.3%; Score 525.4; DB 22; Length 881;  
SQ Best Local Similarity 99.8%; Pred. No. 1.7e-120;  
Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX 3363 AGCTGAACCTACTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422  
DB |||||  
XX 3423 CAGAGACTCTGCCAGGCCACTGTGTGACTTGGAGGTGTGGAGCCCTCAATGACAAC 3482  
DB |||||  
XX 415 CAGAGACTCTGCCAGGCCACTGTGTGACTTGGAGGTGTGGAGCCCTCAATGACAAC 474  
DB |||||  
XX 3483 TGCGGCTGACCCAGAACTACAGGCTACAGCCACTTCTGTGTACTTGGTGCCCTCA 3542  
DB |||||  
XX 475 TGGGCTGACCCAGAACTACAGGCTACAGCCACTTCTGTGTACTTGGTGCCCTCA 534  
DB |||||  
XX 3543 ACCGTGAGGCTGCCACTCTGAGCTGCGCGCAGCTGGCCCACTTCTGACCCGCTCC 3602  
DB |||||  
XX 535 ACCGTGAGGCTGCCACTCTGAGCTGCGCGCAGCTGGCCCACTTCTGACCCGCTCC 594  
DB |||||

QY 3603 AGGGCTCTGGGCGAGCATTTGGGGCGTCATGGCAGCTCTGGGCTACCCACTGCCCCAGC 3662  
|||  
DB 595 AGGGCTCTGGGCGAGCATTTGGGGCGTCATGGCAGCTCTGGGCTACCCACTGCCCCAGC 554  
|||  
QY 3663 CGCTGCTGGGACTGAACCCACTTGGACTCTCTGGGCCCTGCCACAGTGAATTCCTCCAGA 3722  
|||  
DB 655 CGCTGCTGGGACTGAACCCACTTGGACTCTCTGGGCCCTGCCACAGTGAATTCCTCCAGA 714  
|||  
QY 3723 AGATGAGAGACTTCTGGGCTGCTGAAGGAGCTGCAGACCTTCTCTGCTGGCCCTGGCCCAAG 3782  
|||  
DB 715 AGATGAGAGACTTCTGGGCTGCTGAAGGAGCTGCAGACCTTCTCTGCTGGCCCTGGCCCAAG 774  
|||  
QY 3783 ACTTCAACCGGCTCAAGAGAGATGCAGCCCTCCAGCAGCTGCAGTCACTCCCTGCACCTGG 3842  
|||  
DB 775 ACTTCAACCGGCTCAAGAGAGATGCAGCCCTCCAGCAGCTGCAGTCACTCCCTGCACCTGG 834  
|||  
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTCCCTTT 3889  
|||  
DB 835 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTCCCTTT 881  
|||

RESULT 10  
AAV47510  
ID AAV47510 standard; cDNA; 797 BP.  
XX AC AAV47510;  
XX DT 09-NOV-1998 (first entry)  
XX DE Human neurotrophic factor NNT-1 cDNA.  
XX KW NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
KW peripheral neuropathy; dystrophy; neural retina degeneration;  
KW common variable immunodeficiency; CVID; selective IGA deficiency;  
KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antisepctic;  
KW therapy; ss.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT CDS 90..767  
FT sig\_peptide /\*tag= a  
FT mat\_peptide /\*tag= b  
FT /\*tag= c  
XX WO9833922-A1.  
XX PD 06-AUG-1998.  
XX PF 02-FEB-1998; 98WO-US02363.  
XX PR 30-JAN-1998; 98US-0016534.  
XX PR 03-FEB-1997; 97US-0792019.  
XX PA (AMGE-) AMGEN INC.  
XX PI Chang M, Elliot GS, Sarmiento U, Senaldi G;  
XX WPI; 1998-437475/37.  
XX DR P-PSDB; AAW29715.  
XX  
XX Newly isolated nucleic acid encoding human or murine neurotrophic  
PT factor NNT-1 - useful for treatment of neurological and  
PT immunological diseases or inflammation, also as vaccine adjuvant  
XX Claim 3; Fig 1; 120pp; English.  
XX PS This newly isolated human cDNA sequence (deposited at ATCC 98295)  
CC

CC codes for a novel neurotrophic factor, designated NNT-1 (see  
CC AAV29715), that is a growth factor for neurons and for B or T cells.  
CC It was obtained from a T-cell lymphoma cDNA library by expressed  
CC sequence tag analysis on the basis of homology to CNTF. The  
CC isolated NNT-1 cDNA was used as probe to isolated NNT-1 genomic  
CC DNA (see AAV47511). Vectors containing the cDNA or genomic DNA and  
CC host cells are provided for use in the production of NNT-1  
CC polypeptides. These are used to treat: (i) neurological or  
CC immunological diseases, specifically Alzheimer's, Parkinson's or  
CC Huntington's diseases, amyotrophic lateral sclerosis,  
CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and  
CC degeneration of the neural retina, or conditions characterised by T  
CC or B cell defects, e.g. common variable immunodeficiency (CVID),  
CC selective IgA deficiency, hypogammaglobulinaemia and X-linked  
CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)  
CC inflammation. NNT-1 is also able to boost immunoreactivity and  
CC antibody production following vaccination, and, since it inhibits  
CC tumour necrosis factor production, it may also be useful for  
CC treating sepsis. NNT-1 nucleic acid fragments are also used as  
CC hybridisation probes in diagnostic assays. In addition, cells that  
CC have been engineered to express NNT-1 can be implanted, or nucleic  
CC acids are delivered in gene therapy vectors.  
XX  
SQ Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;  
Query Match 10.3%; Score 523.4; DB 19; Length 797;  
Best Local Similarity 99.8%; Pred. No. 5.1e-120;  
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422  
Db 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 330  
QY 3423 CAGAGACTTGCAGGGGCGACTGTGACTTGAGGTGGGAGGCTCAATGACAAAC 3482  
Db 331 CAGAGACTTGCAGGGGCGACTGTGACTTGAGGTGGGAGGCTCAATGACAAAC 390  
QY 3483 TCGGCTGACCCAGAACTACGAGGCTTACAGCCCTTCTGTACTTCCGTGGCTCA 3542  
Db 391 TCGGCTGACCCAGAACTACGAGGCTTACAGCCCTTCTGTACTTCCGTGGCTCA 450  
QY 3543 ACCGTGAGGCTGACCTGTGAGTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTCC 3602  
Db 451 ACCGTGAGGCTGACCTGTGAGTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTCC 510  
QY 3603 AGGGCTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTCC 3662  
Db 511 AGGGCTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTCC 570  
QY 3663 CGCTGCTGGGACTGAACCCACTTGGACTTCTGGCCCTGCCCAGTACTTCTCCAGA 3722  
Db 571 CGCTGCTGGGACTGAACCCACTTGGACTTCTGGCCCTGCCCAGTACTTCTCCAGA 630  
QY 3723 AGATGAGCACTTCTGGCTGCTGAAGAGAGTGCAGACCTGGCTGTGGCGTGGCCCAAG 3782  
Db 631 AGATGAGCACTTCTGGCTGCTGAAGAGAGTGCAGACCTGGCTGTGGCGTGGCCCAAG 690  
QY 3783 ACTTCAACGGCTCAAGAGAAAGATGAGGCTTCCAGAGTGCAGTACCCCTGCACCTGG 3842  
Db 691 ACTTCAACGGCTCAAGAGAAAGATGAGGCTTCCAGAGTGCAGTACCCCTGCACCTGG 750  
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTTCTGCTTCTGCT 3887  
Db 751 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTTCTGCTTCTGCT 795  
RESULT 11  
AAV22652  
ID AAV22652 standard; cDNA; 797 BP.  
XX  
AC AAV22652;  
XX  
DT 13-JUL-1998 (first entry)

XX cDNA encoding human neurotrophic factor NNT-1.  
DE Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;  
XX treatment; neurological disease; degeneration; Parkinson's disease;  
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke; ss.  
KW  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
PH CDS 90..767  
FT /\*tag= a  
FT sig\_peptide 90..170  
FT /\*tag= b  
FT mat\_peptide 171..764  
FT /\*tag= c  
XX  
XX US5741772-A.  
PN  
XX 21-APR-1998.  
PD  
XX 03-FEB-1997; 97US-0792019.  
PF  
XX 03-FEB-1997; 97US-0792019.  
PR  
XX (AMGE-) AMGEN INC.  
PA  
XX Chang M;  
PI  
XX WPI; 1998-260526/23.  
DR P-PSDB; AAV56141.  
XX  
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -  
PT useful for stimulating growth of motor and sympathetic neurons  
XX  
XX Disclosure; Fig 1; 41pp; English.  
XX  
XX The present sequence encodes a human neurotrophic factor, designated  
CC NNT-1, which is capable of stimulating growth of motor or sympathetic  
CC neurons. The NNT-1 protein is useful in the treatment of neurological  
CC diseases characterised by the degeneration and death of particular  
CC classes of neurons. These diseases specifically include Parkinson's  
CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,  
CC stroke and various degenerative disorders affecting vision.  
XX  
SQ Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;  
Query Match 10.3%; Score 523.4; DB 19; Length 797;  
Best Local Similarity 99.8%; Pred. No. 5.1e-120;  
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422  
Db 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 330  
QY 3423 CAGAGACTTGCAGGGGCGACTGTGACTTGAGGTGGGAGGCTCAATGACAAAC 3482  
Db 331 CAGAGACTTGCAGGGGCGACTGTGACTTGAGGTGGGAGGCTCAATGACAAAC 390  
QY 3483 TCGGCTGACCCAGAACTACGAGGCTTACAGCCCTTCTGTACTTCCGTGGCTCA 3542  
Db 391 TCGGCTGACCCAGAACTACGAGGCTTACAGCCCTTCTGTACTTCCGTGGCTCA 450  
QY 3543 ACCGTGAGGCTGACCTGTGAGTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTCC 3602  
Db 451 ACCGTGAGGCTGACCTGTGAGTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTCC 510  
QY 3603 AGGGCTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTCC 3662  
Db 511 AGGGCTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTTGTGAGTGTGGGCGAGCTCC 570  
QY 3663 CGCTGCTGGGACTGAACCCACTTGGACTTCTGGCCCTGCCCAGTACTTCTCCAGA 3722

Db 571 CGTGCCTGGAGTGAACCCACTTGGACTCTCGGCCCTGGCCCAAGTGAAGTCTTCCAG 630  
 QY 3723 AGATGACGACTTCTGGCTGTGAAGAGCTGACAGCTGCTGTGGGCTCGGCCAAGG 3782  
 Db 631 AGATGACGACTTCTGGCTGTGAAGAGCTGACAGCTGCTGTGGGCTCGGCCAAGG 690  
 QY 3783 ACTTCAACCGGCTCAAGAGAGATGACGCTTCCAGCAGCTGCAAGTCAACCTGCACTGG 3842  
 Db 691 ACTTCAACCGGCTCAAGAGAGATGACGCTTCCAGCAGCTGCAAGTCAACCTGCACTGG 750  
 QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCTGCTCCCTCC 3887  
 Db 751 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCTGCTCCCTCC 795

RESULT 12  
 AAA39481  
 ID AAA39481 standard; cDNA; 797 BP.  
 XX AAA39481;  
 AC AAA39481;  
 DT 24-AUG-2000 (first entry)  
 XX Human NNT-1 cDNA.  
 XX NNT-1; human; neurotrophic factor; neuroprotective; treatment;  
 KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;  
 KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
 KW Huntington's disease; peripheral neuropathy; neural retina degeneration;  
 KW retinopathy; immune disorder; hematopoietic disorder; ss.  
 XX Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 90..767  
 FT /\*tag= a  
 FT /product= "NNT-1"  
 XX  
 XX US6054294-A.  
 XX  
 XX 25-APR-2000.  
 XX  
 XX 12-DEC-1997; 97US-0988819.  
 XX  
 XX 03-FEB-1997; 97US-0792019.  
 XX  
 XX (AMGE-) AMGEN INC.  
 XX  
 XX Chang M;  
 XX  
 XX WPI; 2000-338492/29.  
 XX P-PSDB; AAY87813.  
 XX  
 XX New nucleic acids encoding neurotrophic factors useful for stimulating  
 XX growth of motor or sympathetic neurons for treating neuron cell damage  
 XX  
 XX Claim 1a; Fig 1; 42pp; English.  
 XX  
 XX This invention describes a novel nucleic acid molecule (I) encoding a  
 XX novel neurotrophic factor (NNT-1) (II) which has neurotrophic,  
 XX neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and  
 XX ophthalmological activity. (I) is useful for producing NNT-1  
 XX polypeptides which are useful for treating patients in whom various  
 XX cells of the central, autonomic, or peripheral nervous system have  
 XX degenerated and/or have been damaged by congenital disease, trauma,  
 XX mechanical damage, surgery, stroke, ischemia, infection, metabolic  
 XX disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1  
 XX proteins are used to treat diseases like Alzheimer's, Parkinson's,  
 XX amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's  
 XX disease, peripheral neuropathy induced by diabetes or other metabolic  
 XX disorders, and/or dystrophies or degeneration of the neural retina such

CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of  
 CC night blindness, progressive cone-rod degeneration, immune disorders and  
 CC hematopoietic disorders. (I) is effective in treating neurological  
 CC conditions and promotes neuron regeneration. Neural functions are  
 CC effectively restored in patients suffering from various neurological  
 CC disorders. This sequence encodes the human NNT-1 protein described in the  
 CC method of the invention.

SQ Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;

Query Match 10.3%; Score 523.4; DB 21; Length 797;  
 Best Local Similarity 99.8%; Pred. No. 5.1e-120;  
 Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422  
 Db 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 330  
 QY 3423 CAGAGACTTGTCCAGGGCCACTTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAAAC 3482  
 Db 331 CAGAGACTTGTCCAGGGCCACTTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAAAC 390  
 QY 3483 TGGGCTGACCCAGAACTACAGGCCCTACAGCCACTTCTGTGTTACTTGGTGGCTTCA 3542  
 Db 391 TGGGCTGACCCAGAACTACAGGCCCTACAGCCACTTCTGTGTTACTTGGTGGCTTCA 450  
 QY 3543 ACCGTGAGGCTGCCACTGTGAGCTGCGCGCAGCCCTGGCCACTTCTGACACGACCTCC 3602  
 Db 451 ACCGTGAGGCTGCCACTGTGAGCTGCGCGCAGCCCTGGCCACTTCTGACACGACCTCC 510  
 QY 3603 AGGCTCTGTGGGCGAGCATTTGCGGGCGTATGCGAGCTCTGGGCTACCCACTGCCCCAGC 3662  
 Db 511 AGGCTCTGTGGGCGAGCATTTGCGGGCGTATGCGAGCTCTGGGCTACCCACTGCCCCAGC 570  
 QY 3663 CGTGCCTGGGAGTGAACCCACTTGGACTCTCTGGCCCTGGCCCAAGTGAATCTTCCAGA 3722  
 Db 571 CGTGCCTGGGAGTGAACCCACTTGGACTCTCTGGCCCTGGCCCAAGTGAATCTTCCAGA 630  
 QY 3723 AGATGACGACTTCTGGCTGTGAAGAGCTGACAGCTGCTGGCTGGGCTCGGCCAAGG 3782  
 Db 631 AGATGACGACTTCTGGCTGTGAAGAGCTGACAGCTGCTGGCTGGGCTCGGCCAAGG 690  
 QY 3783 ACTTCAACCGGCTCAAGAGAGATGACGCTTCCAGCAGCTGCAAGTCAACCTGCACTGG 3842  
 Db 691 ACTTCAACCGGCTCAAGAGAGATGACGCTTCCAGCAGCTGCAAGTCAACCTGCACTGG 750  
 QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCTGCTCCCTCC 3887  
 Db 751 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCTGCTCCCTCC 795

RESULT 13

ABK11647  
 ID ABK11647 standard; cDNA; 797 BP.  
 XX  
 XX AC  
 XX ABK11647;  
 XX  
 XX 05-JUN-2002 (first entry)  
 XX  
 XX Human cDNA encoding novel neurotrophic factor NNT1.

DE Human cDNA encoding novel neurotrophic factor NNT1.  
 XX Human; ss; gene; NNT1; neurotrophic factor; IgE-related disease;  
 KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;  
 KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;  
 KW vascular restenosis; rheumatoid arthritis; psoriatic arthritis;  
 KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;  
 KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;  
 KW inflammatory bowel disease; transplant rejection; reproductive disorder;  
 KW graft versus host disease; infertility; miscarriage; preterm labour.  
 XX Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FH

CDS 90..767  
 /\*tag= a  
 /product= "NNT1"  
 WO200215977-A2.  
 28-FEB-2002.  
 17-AUG-2001; 2001WO-US25906.  
 18-AUG-2000; 2000US-226436P.  
 16-AUG-2001; 2001US-0931704.  
 (AMGE-) ANGEN INC.  
 Senaldi G;  
 WPI; 2002-280867/32.  
 P-PSDB; AAU78176.  
 Treating Immunoglobulin E-related disease, modulating IgE levels in a patient, preventing IgE-related disease and treating allergic diseases, involves administering NNT-1 inhibitor to a patient -  
 Claim 2; Fig 1; 63pp; English.  
 The invention relates to treating Immunoglobulin E (IgE)-related disease, modulating IgE levels in a patient, preventing an IgE-related disease, and treating allergic diseases, comprising administering a therapeutically effective amount of novel neurotrophic factor (NNT)-1 inhibitor to a patient. Also included are a method of diagnosing an IgE-related disease or susceptibility to an IgE-related disease, by determining the presence or amount of expression of an NNT1 polypeptide encoded by a NNT1 nucleotide sequence, its fragment or naturally occurring variant, and diagnosing an IgE-related disease or susceptibility of an IgE-related disease based on the presence or amount of expression of the polypeptide and a pharmaceutical composition for use in treating IgE-related disease, comprising the NNT1 inhibitor.  
 The NNT1 inhibitor is useful for preventing and treating IgE-related disease, modulating IgE levels, and treating allergic diseases e.g. Type I allergic disease, allergic rhinitis, eczema, dermatitis, pollinosis, asthma, immune diseases and disorders, diseases involving abnormal cell proliferation including cancer, arteriosclerosis and vascular restenosis, diseases and conditions relating to dysfunction of immune system including rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, osteoarthritis, inflammatory joint disease, autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory bowel disease, transplant rejection, and graft versus host disease, and reproductive diseases and disorders including infertility, miscarriage, preterm labour and delivery, and endometriosis. The present sequence encodes human NNT1.  
 Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;  
 Query Match 10.3%; Score 523.4; DB 24; Length 797;  
 Best Local Similarity 99.8%; Pred. No. 5,1e-120;  
 Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 3422  
 DB 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 330  
 QY 3423 CAGAGACTGTGCCAGGGCCACCTGTTGACCTGGAGGTGTGCGAAGCTCAATGACAAC 3482  
 DB 331 CAGAGACTGTGCCAGGGCCACCTGTTGACCTGGAGGTGTGCGAAGCTCAATGACAAC 390  
 QY 3483 TGGCGGTGACCCAGAACTACGAGGCCCTTACAGCCACCTTCTGTGTTACTTGGTGGCCCTCA 3542  
 DB 391 TGGCGGTGACCCAGAACTACGAGGCCCTTACAGCCACCTTCTGTGTTACTTGGTGGCCCTCA 450  
 QY 3543 ACCGTGAGGTGACCTGCTGAGCTGGCGCGAGCTGGCCACCTTCTGACGACGACCTCC 3602  
 DB 451 ACCGTGAGGTGACCTGCTGAGCTGGCGCGAGCTGGCCACCTTCTGACGACGACCTCC 510

QY 3603 AGGCGCTGCTGGGCGAGCATTTGGGGCGTCTATGGCAGCTCTGGGCTACCCACTGCCCCAGC 3662  
 DB 511 AGGCGCTGCTGGGCGAGCATTTGGGGCGTCTATGGCAGCTCTGGGCTACCCACTGCCCCAGC 570  
 QY 3663 CGTGGCTGGGACTGAACCCACTTGGACTCTGCGCCCTGCCAGTCACTTCTCCAGA 3722  
 DB 571 CGTGGCTGGGACTGAACCCACTTGGACTCTGCGCCCTGCCAGTCACTTCTCCAGA 630  
 QY 3723 AGATGGACGACTTCTGGCTGCTGAAGAGCTGCAGACTGCTGGCGCTGGGCAAGG 3782  
 DB 631 AGATGGACGACTTCTGGCTGCTGAAGAGCTGCAGACTGCTGGCGCTGGGCAAGG 690  
 QY 3783 ACTTCAACCGGCTCAAGAAGAAGATGCAGCCTCCAGCAGCTGCAGTCACTTCTGCTCC 3842  
 DB 691 ACTTCAACCGGCTCAAGAAGAAGATGCAGCCTCCAGCAGCTGCAGTCACTTCTGCTCC 750  
 QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCTGCTCCGCC 3887  
 DB 751 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCTGCTCCGCC 795  
 RESULT 14  
 ID ABA09140  
 ID ABA09140 standard; cDNA; 968 BP.  
 XX ABA09140;  
 AC ABA09140;  
 DT 11-JAN-2002 (first entry)  
 XX Human cardiostrophin-like cytokine homologue cDNA, SEQ ID NO:916.  
 DE Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;  
 KW cytotatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cytotatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnery; antiulcer; ss.  
 XX Homo sapiens.  
 OS  
 PN WC200157188-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US03800.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI; 2001-457740/49.  
 DR P-PSDB; ABB11896.  
 XX  
 PT Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 PS Claim 1; Page 793-794; 1963pp; English.  
 XX  
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a

nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoietic regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a cDNA encoding a novel human polypeptide of the invention.

SQ Sequence 968 BP; 179 A; 353 C; 244 G; 192 T; 0 other;

Query Match	9.7%	Score 495.4	DB 22	Length 968
Best Local Similarity	99.8%	Pred. No. 5.4e-113		
Matches 496	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	3363	AGCTGMACTACTCGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGGG	3422	
Db	472	ATCTGAACTACTCGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGGG	531	
QY	3423	CAGAGACTCTGCCAGGGCCACTGTGTGACTTTGGAGGTGTGCGAAGCCTCAATGACAAAC	3482	
Db	532	CAGAGACTCTGCCAGGGCCACTGTGTGACTTTGGAGGTGTGCGAAGCCTCAATGACAAAC	591	
QY	3483	TGCGGCTGACCCAGAACTACGAGGCCCTACAGCCACCTTCTGTGTACTTGGTGGCCCTCA	3542	
Db	592	TGCGGCTGACCCAGAACTACGAGGCCCTACAGCCACCTTCTGTGTACTTGGTGGCCCTCA	651	
QY	3543	ACCGTCAGGCTGCCACTGCTGAGCTGCGCCGAGCGCTGGGCCACTTCTGTGACCAAGCCTCC	3602	
Db	652	ACCGTCAGGCTGCCACTGCTGAGCTGCGCCGAGCGCTGGGCCACTTCTGTGACCAAGCCTCC	711	
QY	3603	AGGCGCTGCTGGGAGAGCATTCGGGCGGTCATGGCAGCTCTGGGCTACCCACTGCCCCAGC	3662	
Db	712	AGGCGCTGCTGGGAGAGCATTCGGGCGGTCATGGCAGCTCTGGGCTACCCACTGCCCCAGC	771	
QY	3663	CGCTGCGCTGGGACTGAAACCCACTTCCTGGCCCTCGCCCAACAGTGACTTCTCTCAGA	3722	
Db	772	CGCTGCGCTGGGACTGAAACCCACTTCCTGGCCCTCGCCCAACAGTGACTTCTCTCAGA	831	
QY	3723	AGATGAGACGACTTCTGGCTGTGTGAAGGAGCTGCAGACTGGCTGTGGCGCTCGGCCAAGG	3782	
Db	832	AGATGAGACGACTTCTGGCTGTGTGAAGGAGCTGCAGACTGGCTGTGGCGCTCGGCCAAGG	891	
QY	3783	ACTTCAACCGGCTCAAGAGAGAGATGAGCCCTCAGACGCTGCAGTCAACCTGCACCTGG	3842	
Db	892	ACTTCAACCGGCTCAAGAGAGAGATGAGCCCTCAGACGCTGCAGTCAACCTGCACCTGG	951	

QY 3843 GGGCTCATGGCTTCTGA 3859  
Db 952 GGGCTCATGGCTTCTGA 968

RESULT 15

AAK52532  
ID AAK52532 standard; cDNA; 968 BP.

AAK52532;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 2061.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW  
KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200157190-A2.

PD 09-AUG-2001.

05-FEB-2001: 2001WO-US04098.

PR 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-06933325.  
PR 30-NOV-2000; 2000US-0728422

FK XX  
30-NOV-2000; 2000US-0/284ZZ.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;

DR WPI: 2001-476283/51.

DR P-PSDB; AAM79399.

PT Nucleic acids encoding polypeptides with cytokine-like activities

PT useful in diagnosis and gene therapy -

PS Claim 1: Page 4448-4448: 6221pp: English:

The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM8020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

Sequence 968 BP: 179 A: 353 C: 244 G: 192 T: 0 other:

Query Match 9.7%; Score 495.4; DB 22;  
Best Local Similarity 99.8%; Pred. No. 5.4e-113;  
Matches 496; Conservative 0; Mismatches 1; Indels 0;

```
Db 472 ATCTGAACCTACCTGGGCCCTTTCAACGAGCAGACTTCAACCCCTCCCGCTGGGG 531
Qy 3423 CAGAGACTCTGCCAGGGCCACTGTGTGACTTGGAGGTGTGGGAGCCCTCAATGACAAAC 3482
Db 532 CAGAGACTCTGCCAGGGCCACTGTGTGACTTGGAGGTGTGGGAGCCCTCAATGACAAAC 591
Qy 3483 TGGGCTGACCCAGAACCTACGAGGCTTACAGCCACTTCTGTGTTACTTGGTGGGCTCA 3542
Db 592 TGGGCTGACCCAGAACCTACGAGGCTTACAGCCACTTCTGTGTTACTTGGTGGGCTCA 651
Qy 3543 ACCGTGAGGCTGCCACTGTGTGAGCTGCGCCGAGCCTTGCCCACTTCTGACCCAGCTCC 3602
Db 652 ACCGTGAGGCTGCCACTGTGTGAGCTGCGCCGAGCCTTGCCCACTTCTGACCCAGCTCC 711
Qy 3603 AGGGCTGCTGGGCGAGCATTTGGGGCTCATGGCAGCTTGGGCTACCCACTGCCCCAGC 3662
Db 712 AGGGCTGCTGGGCGAGCATTTGGGGCTCATGGCAGCTTGGGCTACCCACTGCCCCAGC 771
Qy 3663 CGCTGCCCTGGGACTGAACCCACTTGGACTCCTGGGCCCTGCCACAGTACTTCTCCAGA 3722
Db 772 CGCTGCCCTGGGACTGAACCCACTTGGACTCCTGGGCCCTGCCACAGTACTTCTCCAGA 831
Qy 3723 AGATGGAAGACTTCTGGCTGTGAAGAGCTGCAGACCTGGGTGTGGCGCTCGGCCAAGG 3782
Db 832 AGATGGAAGACTTCTGGCTGTGAAGAGCTGCAGACCTGGGTGTGGCGCTCGGCCAAGG 891
Qy 3783 ACTTCAACCGGCTCAAGAGAGATGCAGCCTCCAGCAGCTGCAGTCAACCTGCACCTGG 3842
Db 892 ACTTCAACCGGCTCAAGAGAGATGCAGCCTCCAGCAGCTGCAGTCAACCTGCACCTGG 951
Qy 3843 GGGCTCATGGCTTCTGA 3859
Db 952 GGGCTCATGGCTTCTGA 968
```

Search completed: February 1, 2003, 09:12:09  
Job time : 1346.47 secs

1	5087	100.0	5087	10	US-09-931-704-3	Sequence 3, Appl	
2	523.4	10.3	797	10	US-09-931-704-1	Sequence 1, Appl	
3	493.4	9.7	495	10	US-09-864-761-23175	Sequence 21375, A	
C	492	9.7	492	10	US-09-864-761-6462	Sequence 6462, Ap	
5	437.2	8.6	465	10	US-09-867-701-4667	Sequence 4667, Ap	
6	429.6	8.4	419	10	US-09-931-704-4	Sequence 4, Appl	
C	7	43.9	8.4	819	10	US-09-867-701-4636	Sequence 4636, Ap
C	8	358.8	7.0	360	10	US-09-867-701-2610	Sequence 2610, Ap
9	250.6	4.9	269	10	US-09-867-701-5382	Sequence 5382, Ap	
10	192.8	3.8	231	10	US-09-867-701-2639	Sequence 2639, Ap	
11	165.2	3.2	283	10	US-09-765-231A-75	Sequence 75, Appl	
12	57.6	1.1	53522	9	US-09-904-968A-1	Sequence 1, Appl	
13	54.4	1.1	1064	10	US-09-804-682-29	Sequence 29, Appl	
14	53	1.0	495	10	US-09-864-761-10689	Sequence 10689, A	
15	53	1.0	1065	10	US-09-804-682-33	Sequence 33, Appl	
16	53	1.0	659158	9	US-09-771-208-20	Sequence 20, Appl	
17	52.4	1.0	267	10	US-09-864-761-27984	Sequence 27984, A	
18	52.4	1.0	474	10	US-09-864-761-11284	Sequence 11284, A	
19	52.4	1.0	3809	12	US-10-001-870-68	Sequence 68, Appl	

Db	181	CGCCCCATCTGATACACTTAACCGGACCAAGTCACAGCCCTCCAACCTCACCCCTCGCTGCTGCC	240
Qy	241	CAGAGCTCACCAATCTCTTTGTGGACTCAAACTCAACCGCACTAAATCAACCAAAATCCCA	300
Db	241	CAGAGCTCACCAATCTCTTTGTGGACTCAAACTCAACCGCACTAAATCAACCAAAATCCCA	300
Qy	301	AGTCTAAACTAATCTGAAACCTTTTAAAGTAACCCAGTCTCTTAAACCTTAACCTTAGCCCAAT	360
Db	301	AGTCTAAACTAATCTGAAACCTTTTAAAGTAACCCAGTCTCTTAAACCTTAACCTTAGCCCAAT	360
Qy	361	GCCAAATTAATCTACCCCTAGCCAAACCCCTAACTGCGCTTTGCCAGTCCCAAAGTGTCCACTG	420
Db	361	GCCAAATTAATCTACCCCTAGCCAAACCCCTAACTGCGCTTTGCCAGTCCCAAAGTGTCCACTG	420
Qy	421	AATCTCACTTGTGCTCACTGTGAAAATCCAGAAAAGCATATTTTCCCACTGCCACAT	480
Db	421	AATCTCACTTGTGCTCACTGTGAAAATCCAGAAAAGCATATTTTCCCACTGCCACAT	480
Qy	481	CCCTCTTTACAGCACCCCAACCTGCGCTCTGACCTCTGGATCTCTGGATCTGCAAACT	540
Db	481	CCCTCTTTACAGCACCCCAACCTGCGCTCTGACCTCTGGATCTCTGGATCTGCAAACT	540
Qy	541	CTGAGTGCATCAGCCAAACAGCCGACTCGTCAAAATGCACTCTCTCCCTTCCTGTGCC	600
Db	541	CTGAGTGCATCAGCCAAACAGCCGACTCGTCAAAATGCACTCTCTCCCTTCCTGTGCC	600
Qy	601	CCACCTTTGAGGCTGATGGAAAGGCTCATTTGAAGTCCAACTTTTCCCACTAACACC	660
Db	601	CCACCTTTGAGGCTGATGGAAAGGCTCATTTGAAGTCCAACTTTTCCCACTAACACC	660
Qy	661	AGAAACGGGGTGAACCTCCACACTGCGACCTGAGAGTGAGCACTAAATCTCTCT	720
Db	661	AGAAACGGGGTGAACCTCCACACTGCGACCTGAGAGTGAGCACTAAATCTCTCT	720
Qy	721	CAATCTAAACCCCACTCACTTTCCACACTCAGGAATCACATCTCCTAGATATACCCAAA	780
Db	721	CAATCTAAACCCCACTCACTTTCCACACTCAGGAATCACATCTCCTAGATATACCCAAA	780
Qy	781	ACTAAGCCCAATAAGGACGCCGACCCCTAGTGTCTAAACCTTAACCTTGTCTCTATGG	840
Db	781	ACTAAGCCCAATAAGGACGCCGACCCCTAGTGTCTAAACCTTAACCTTGTCTCTATGG	840
Qy	841	GTGAGTCTGTTCTTGGCGCGCGCTCTCTGCTTCTCCCTTAGAGCTGAGCTGCTC	900
Db	841	GTGAGTCTGTTCTTGGCGCGCGCTCTCTGCTTCTCCCTTAGAGCTGAGCTGCTC	900
Qy	901	AGCCTGCCAGCTCTGACATGTGTCTGCCACCTCTGACTCCCTCAAGCTGCAGTGG	960
Db	901	AGCCTGCCAGCTCTGACATGTGTCTGCCACCTCTGACTCCCTCAAGCTGCAGTGG	960
Qy	961	GACTGAAAGACTGGCAGGAGCTTAGGGTAACTGGAACACAGGAGGTGCAGCTGCAGT	1020
Db	961	GACTGAAAGACTGGCAGGAGCTTAGGGTAACTGGAACACAGGAGGTGCAGCTGCAGT	1020
Qy	1021	CCCTAGGCTTGGCCCGCTCCCTCCATGTACACATATATATGTGGCACACACAGTG	1080
Db	1021	CCCTAGGCTTGGCCCGCTCCCTCCATGTACACATATATATGTGGCACACACAGTG	1080
Qy	1081	GCACATGTCGAAAGACTCTCTCAGCTGACACACAGATCAATCTCAAGTATCTACTGAT	1140
Db	1081	GCACATGTCGAAAGACTCTCTCAGCTGACACACAGATCAATCTCAAGTATCTACTGAT	1140
Qy	1141	AGACATCTATGCTGCGCAAGTCTCATCTCTCAAAACATACACATGCTCTCTCTCCC	1200
Db	1141	AGACATCTATGCTGCGCAAGTCTCATCTCTCAAAACATACACATGCTCTCTCTCCC	1200
Qy	1201	GTCTTGGCAGGAGTGTCTCCCTCTCCATCTCCCTCTGCTCCATCTGCTGCTGCTCCACCC	1260
Db	1201	GTCTTGGCAGGAGTGTCTCCCTCTCCATCTCCCTCTGCTCCATCTGCTGCTGCTCCACCC	1260
Qy	1261	TCACCCCCCAACCCAGGCTGGGACAGACACCTGAGGGGTGCGAGCTGCTTCCC	1320



QY 2401 TCCCTGGGCCCCAGCGCTCCCGAGGGTTGGAAAGGGCTCTGCGCTCTTCCCTATACC 2460  
DB 2401 TCCCTGGGCCCCAGCGCTCCCGAGGGTTGGAAAGGGCTCTGCGCTCTTCCCTATACC 2460  
QY 2461 ATGCTGTCTTCATAGCCTTCTCTCTGCTCTACTCATGAGAGCTGCTCCATTTCTCTT 2520  
DB 2461 ATGCTGTCTTCATAGCCTTCTCTCTGCTCTACTCATGAGAGCTGCTCCATTTCTCTT 2520  
QY 2521 CTGCAACCCCTGCTCTATACAGTGAACCCCTTCTTTCGGAGTGTAGTAGTACCCGCTC 2580  
DB 2521 CTGCAACCCCTGCTCTATACAGTGAACCCCTTCTTTCGGAGTGTAGTAGTACCCGCTC 2580  
QY 2581 TCCCGAGCCCTCAGCTGAGTGGGCTGGGTGTGACGGCAAAATGGGGCTCTGTTTCCA 2640  
DB 2581 TCCCGAGCCCTCAGCTGAGTGGGCTGGGTGTGACGGCAAAATGGGGCTCTGTTTCCA 2640  
QY 2641 ATGGGCCACTCTCATCTCTCTCTTCTTCTTGTGTCAGAAAACTTGTGCTTCACTCCA 2700  
DB 2641 ATGGGCCACTCTCATCTCTCTCTTGTGTCAGAAAACTTGTGCTTCACTCCA 2700  
QY 2701 CCCTCTAGTTCGCGACCCCTTCTCTCTGCTTTCCTGCTGCCAAATTTCTCAAGGA 2760  
DB 2701 CCCTCTAGTTCGCGACCCCTTCTCTCTGCTTTCCTGCTGCCAAATTTCTCAAGGA 2760  
QY 2761 GTGCTTACACCCCTCTGCTCTCCACTTCTCTCCACCACTCACTTCTTAAACCCCTGCAA 2820  
DB 2761 GTGCTTACACCCCTCTGCTCTCCACTTCTCTCCACCACTCACTTCTTAAACCCCTGCAA 2820  
QY 2821 TCTGGCTTCAGGCCCCAGCAATGGTCTCTCAGAGTGTGTCAGGACCTCTTGGCCAA 2880  
DB 2821 TCTGGCTTCAGGCCCCAGCAATGGTCTCTCAGAGTGTGTCAGGACCTCTTGGCCAA 2880  
QY 2881 CCCGACAGTGTGTTGAAGGCTCATCTCTCTGCTGCTGTTTTCAGCCACACTGCTGAG 2940  
DB 2881 CCCGACAGTGTGTTGAAGGCTCATCTCTCTGCTGCTGTTTTCAGCCACACTGCTGAG 2940  
QY 2941 CGCTGCTCCCTCTCGAACTCTCTCTCTGCTCTCTGCACTCTCTGCGGCCACCTTCTA 3000  
DB 2941 CGCTGCTCCCTCTCGAACTCTCTCTCTGCTCTCTGCACTCTCTGCGGCCACCTTCTA 3000  
QY 3001 CCTCTCAGCTCTCCAGGCTCTCTCTCTCTCTGCTCTGCTGCCCAAGCGGGCACTCTC 3060  
DB 3001 CCTCTCAGCTCTCCAGGCTCTCTCTCTCTCTGCTCTGCTGCCCAAGCGGGCACTCTC 3060  
QY 3061 CCAAGTTTGGCCACCCAGCCAACTCAGACGCTCTCTCTGAGCGCTCTGTCGCTCTCTC 3120  
DB 3061 CCAAGTTTGGCCACCCAGCCAACTCAGACGCTCTCTCTGAGCGCTCTGTCGCTCTCTC 3120  
QY 3121 CTCCTCTCTTTTCTACGCTCTCCATTTGGAGAGCTCAACCGCCACTGCTTCAACTGTC 3180  
DB 3121 CTCCTCTCTTTTCTACGCTCTCCATTTGGAGAGCTCAACCGCCACTGCTTCAACTGTC 3180  
QY 3181 ACCTGCATACAAATGATATCTTATTTGAAATACTCAGGAGGCGCATGAAACAAAGAGCC 3240  
DB 3181 ACCTGCATACAAATGATATCTTATTTGAAATACTCAGGAGGCGCATGAAACAAAGAGCC 3240  
QY 3241 TAGCATGAGACAGGGCAGTGTGAGGGGACACAAATAAGAACTTTGGGAGCAGGTA 3300  
DB 3241 TAGCATGAGACAGGGCAGTGTGAGGGGACACAAATAAGAACTTTGGGAGCAGGTA 3300  
QY 3301 TCTCTTTGGTGTGAGCAGGGCTCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 3360  
DB 3301 TCTCTTTGGTGTGAGCAGGGCTCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 3360  
QY 3361 ACAGCTGAATCTACCTGGGCCCTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGG 3420  
DB 3361 ACAGCTGAATCTACCTGGGCCCTTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGG 3420  
QY 3421 GGCAGAGACTCTGCCAGGGCACTGTGACTTGGAGTGTGGGAGGCGCTCAATGACAA 3480  
DB 3421 GGCAGAGACTCTGCCAGGGCACTGTGACTTGGAGTGTGGGAGGCGCTCAATGACAA 3480

QY 3481 ACTCGGCTGTACCCAGAACTACGAGGCTCTACGAGCACTCTGTGTATTCTTGCCTGGCT 3540  
DB 3481 ACTCGGCTGTACCCAGAACTACGAGGCTCTACGAGCACTCTGTGTATTCTTGCCTGGCT 3540  
QY 3541 CAAACGCTCAGGCTGCGACTGCTGAGCTGCGCGGAGCGCTGCGCCACTTCTGACCGAGCT 3600  
DB 3541 CAAACGCTCAGGCTGCGACTGCTGAGCTGCGCGGAGCGCTGCGCCACTTCTGACCGAGCT 3600  
QY 3601 CCAGGGCTGCTGCGGAGCACTTTCGGGGCTCATGCGAGCTCTGGGCTACCCACTTGCCTCA 3660  
DB 3601 CCAGGGCTGCTGCGGAGCACTTTCGGGGCTCATGCGAGCTCTGGGCTACCCACTTGCCTCA 3660  
QY 3661 GCCGCTGCTGGGACTGAAACCACTTGGAGCTCTGGGCTGCCCACTGACTTCTCTCCA 3720  
DB 3661 GCCGCTGCTGGGACTGAAACCACTTGGAGCTCTGGGCTGCCCACTGACTTCTCTCCA 3720  
QY 3721 GAAGATGGAAGCACTTCTGGCTGTGGAAGGAGCTGAGACCTGGGCTGTGGGCTCGGCCAA 3780  
DB 3721 GAAGATGGAAGCACTTCTGGCTGTGGAAGGAGCTGAGACCTGGGCTGTGGGCTCGGCCAA 3780  
QY 3781 GGACTTCAAACCGGCTCAAGAAAGATGCGAGCTTCCAGAGCTGCGAGTCACTGCACT 3840  
DB 3781 GGACTTCAAACCGGCTCAAGAAAGATGCGAGCTTCCAGAGCTGCGAGTCACTGCACT 3840  
QY 3841 GGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900  
DB 3841 GGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900  
QY 3901 CCACCTTTGAGAGCCAGCCCTGTATGCGCAACACTTGTGAGCCAGGAGCAGAGCTG 3960  
DB 3901 CCACCTTTGAGAGCCAGCCCTGTATGCGCAACACTTGTGAGCCAGGAGCAGAGCTG 3960  
QY 3961 TGAGCCTCTGSCCTTCTCTGAGCCGCTGGGCTGTGATGCGATCAGCCCTGCTCTCTC 4020  
DB 3961 TGAGCCTCTGSCCTTCTCTGAGCCGCTGGGCTGTGATGCGATCAGCCCTGCTCTCTC 4020  
QY 4021 CCCACCTCCCAAGGCTTACCGAGCTGGGAGGAGTACAGTAGGCGCTCTCTCTCTCTCTG 4080  
DB 4021 CCCACCTCCCAAGGCTTACCGAGCTGGGAGGAGTACAGTAGGCGCTCTCTCTCTCTG 4080  
QY 4081 TTTCTACAGGAAGTCTGCTCGAGGAGTGTGAAGTGTGAGTGTGAGTGTGAGAGGCGCT 4140  
DB 4081 TTTCTACAGGAAGTCTGCTCGAGGAGTGTGAAGTGTGAGTGTGAGTGTGAGAGGCGCT 4140  
QY 4141 CATGSCCTCTGCTCTCTGCTCACTTGTGCGAGTGTGCGCCAGCCAGCCCTCAGGTGCA 4200  
DB 4141 CATGSCCTCTGCTCTCTGCTCACTTGTGCGAGTGTGCGCCAGCCAGCCCTCAGGTGCA 4200  
QY 4201 CATCTGAGGGCAGGGGTTGAGGGGCCACACCAACATGCTCTTCTGGGGTGAAGCCCT 4260  
DB 4201 CATCTGAGGGCAGGGGTTGAGGGGCCACACCAACATGCTCTTCTGGGGTGAAGCCCT 4260  
QY 4261 TTGCTCTGCCCACTCTCTCTTGGATGAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4320  
DB 4261 TTGCTCTGCCCACTCTCTCTTGGATGAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4320  
QY 4321 TCCAAATTCAGGAACAACTGTTGCAATTTCTACAAAGAGATGAGATTAACAGT 4380  
DB 4321 TCCAAATTCAGGAACAACTGTTGCAATTTCTACAAAGAGATGAGATTAACAGT 4380  
QY 4381 CAGGTTGGGGTCTGCACTTGGAGTGGCTTAAACAGAGAGAAATCTGAAAGCAC 4440  
DB 4381 CAGGTTGGGGTCTGCACTTGGAGTGGCTTAAACAGAGAGAAATCTGAAAGCAC 4440  
QY 4441 AGGGCAGGACAGACAGACCCAGAGTCTCCAAAGCAGAGTGTGGGAGGCAACAA 4500  
DB 4441 AGGGCAGGACAGACAGACCCAGAGTCTCCAAAGCAGAGTGTGGGAGGCAACAA 4500  
QY 4501 ACCGAGCTGAGCATCAGGACCTTGCCTCGAATTTGCTTCTCAGTATTACGGTGTCTCTC 4560  
DB 4501 ACCGAGCTGAGCATCAGGACCTTGCCTCGAATTTGCTTCTCAGTATTACGGTGTCTCTC 4560  
QY 4561 TCTGCCCTTTTCCAGGGTATCTGTGGTGTGCGAGGCTGGGGAGGGCAACCATAGCCAC 4620

Db 4561 TCTGCCCTTTCCAGGGTATCTGTGGTGTCCAGGCTGGGAGGCAACCATAGCCAC 4620  
Qy 4621 ACCACAGATTTCTGAAAGTTTCAATGACGTAGATTTGGGGTGTAGGGTGGCAGCT 4680  
Db 4621 ACCACAGATTTCTGAAAGTTTCAATGACGTAGATTTGGGGTGTAGGGTGGCAGCT 4680  
Qy 4681 CCCAAGGCCCTGCCGCCGCCAGCCACCCACTCATGACTCTAAGTGTGTGTATTAATAT 4740  
Db 4681 CCCAAGGCCCTGCCGCCGCCAGCCACCCACTCATGACTCTAAGTGTGTGTATTAATAT 4740  
Qy 4741 TTATTTATTTGGAGATGTTATTTATTTAGATATTTATTTAGAGAAATTTCTATTTCTGTA 4800  
Db 4741 TTATTTATTTGGAGATGTTATTTATTTAGATATTTATTTAGAGAAATTTCTATTTCTGTA 4800  
Qy 4801 TTAACAAATAAATGTTGCCGCCAGAACTTAGTCTCTTGGCCAGCCTCACCCCTCTCTGG 4860  
Db 4801 TTAACAAATAAATGTTGCCGCCAGAACTTAGTCTCTTGGCCAGCCTCACCCCTCTCTGG 4860  
Qy 4861 TGCTCATCAGACTCTTGGCCACCCCTGGCTCCCACTCCCTGCTTGGCTCTGGTGGAGCTGC 4920  
Db 4861 TGCTCATCAGACTCTTGGCCACCCCTGGCTCCCACTCCCTGCTTGGCTCTGGTGGAGCTGC 4920  
Qy 4921 ACAGAGCTCTGGGAAGAGCCCTTTCCTCCCGCACTGGGGCGATGGGGCGACCTCAGA 4980  
Db 4921 ACAGAGCTCTGGGAAGAGCCCTTTCCTCCCGCACTGGGGCGATGGGGCGACCTCAGA 4980  
Qy 4981 CTTACCACTGCTGTGCCACCAACCAACCCCTTGATCCCTCAGTCTCCACACAGCTTC 5040  
Db 4981 CTTACCACTGCTGTGCCACCAACCAACCCCTTGATCCCTCAGTCTCCACACAGCTTC 5040  
Qy 5041 TGTCACCCAGGTTTCCCTCACCCACCTTTGCTAAGTCTTCCCTCA 5087  
Db 5041 TGTCACCCAGGTTTCCCTCACCCACCTTTGCTAAGTCTTCCCTCA 5087

RESULT 2  
US-09-931-704-1  
; Sequence 1, Application US/09931704  
; Patent No. US20020041873A1  
; GENERAL INFORMATION:  
; APPLICANT: Senaldi, Giorgio  
; TITLE OF INVENTION: Methods and Compositions for Treating Ige-Related Disease Using N  
; FILE REFERENCE: A-695  
; CURRENT APPLICATION NUMBER: US/09/931,704  
; PRIOR FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: US 60/226,436  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 797  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (90)..(764)  
; OTHER INFORMATION:  
; NAME/KEY: mat\_peptide  
; LOCATION: (171)..()  
; OTHER INFORMATION: Met at -27  
US-09-931-704-1

Query Match 10.3%; Score 523.4; DB 10; Length 797;  
Best Local Similarity 99.8%; Pred. No. 2.1e-130;  
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 3363 AGCTGAACCTACCTGGGCCCCCTTTTCAACAGGCGACAGCTTCAACCCCTCCCGCTGGGG 3422  
Db 271 ATCTGAACCTACCTGGGCCCCCTTTTCAACAGGCGACAGCTTCAACCCCTCCCGCTGGGG 330  
Qy 3423 CAGAGACTCTGCCAGGGCCACTGTTGACTTGAGGTGTGGGAGCCCTCAATGACAAC 3482

Db 331 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCCTCAATGACAAC 390  
Qy 3483 TGGGGCTGAGCCAGAACTACAGGSCCTACAGCCACCTTCTGTGTACTTGGTGGCCTCA 3542  
Db 391 TGGGGCTGAGCCAGAACTACAGGSCCTACAGCCACCTTCTGTGTACTTGGTGGCCTCA 450  
Qy 3543 ACCTCAGGCTGCCACTGTGAGCTGCCGCCAGCCTGGCCCACTTTGCAACCAAGCTCC 3602  
Db 451 ACCTCAGGCTGCCACTGTGAGCTGCCGCCAGCCTGGCCCACTTTGCAACCAAGCTCC 510  
Qy 3603 AGGGCTGTGGGAGAGATTTGGGGGCTCATGGAGCTCTGGGTACCCACTGCCCGCAGC 3662  
Db 511 AGGGCTGTGGGAGAGATTTGGGGGCTCATGGAGCTCTGGGTACCCACTGCCCGCAGC 570  
Qy 3663 CGCTGCTGGGACTTGAACCCACTTGGAGCTCTGCGCCCTGCCCACTGACTTCTCTCCAGA 3722  
Db 571 CGCTGCTGGGACTTGAACCCACTTGGAGCTCTGCGCCCTGCCCACTGACTTCTCTCCAGA 530  
Qy 3723 AGATGACGACTTCTTGCTGCTGAAGGAGCTGCAGACTGTGCGCTGGGCGCAAGG 3782  
Db 631 AGATGACGACTTCTTGCTGCTGAAGGAGCTGCAGACTGTGCGCTGGGCGCAAGG 690  
Qy 3783 ACTTCAACCGCTCAAGAAGATGAGCCCTCCAGAGCTGCAGTCAACCTTGCACCTGG 3842  
Db 691 ACTTCAACCGCTCAAGAAGATGAGCCCTCCAGAGCTGCAGTCAACCTTGCACCTGG 750  
Qy 3843 GGGCTCATGCTCTGACTTCTGACTTCTGACTTCTCTCTCTTGGCTCCCCC 3887  
Db 751 GGGCTCATGCTCTGACTTCTGACTTCTGACTTCTCTCTCTTGGCTCCCCC 795

RESULT 3  
US-09-864-761-23175/c  
; Sequence 23175, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 23175  
;; LENGTH: 495  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC005849.1  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5  
;; OTHER INFORMATION: SWISSPROT HIT: G63086, EVALU8 8.00e-03  
;; OTHER INFORMATION: NT HIT: G11439486, EVALU8 0.00e+00  
;; OTHER INFORMATION: EST\_HUMAN HIT: A1752561.1, EVALU8 0.00e+00  
US-09-864-761-23175

Query Match 9.7%; Score 493.4; DB 10; Length 495;  
Best Local Similarity 99.8%; Pred. No. 1.7e-122;  
Matches 494; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3365 CTGAACCTCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGGCA 3424  
Db 495 CTGAACCTCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGGCA 436  
QY 3425 GAGACTCTGCCAGGGCCACTGTGAGCTTGAGGTTGCGAAGCTCAATGACAACTG 3484  
Db 435 AAGACTCTGCCAGGGCCACTGTGAGCTTGAGGTTGCGAAGCTCAATGACAACTG 376  
QY 3485 CGGCTGACCCAGAACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGGCTTCAAC 3544  
Db 375 CGGCTGACCCAGAACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGGCTTCAAC 316  
QY 3545 CTTAGGCTGACCTGCTGAGCTGCGCCGCGAGCTTCTGACCAAGCTTCCAG 3604  
Db 315 CTTAGGCTGACCTGCTGAGCTGCGCCGCGAGCTTCTGACCAAGCTTCCAG 256  
QY 3605 GGCCTGCTGGGCGAGCTTGGGGCGCTCATGGGAGCTTGGGCTACCCACTGCCCCAGCG 3664  
Db 255 GGCCTGCTGGGCGAGCTTGGGGCGCTCATGGGAGCTTGGGCTACCCACTGCCCCAGCG 196  
QY 3665 CTGCTGGGAGCTGAACCCACTTGGAGCTCTGGCCCTGCCACAGTCTTCTCCAGAG 3724  
Db 195 CTGCTGGGAGCTGAACCCACTTGGAGCTCTGGCCCTGCCACAGTCTTCTCCAGAG 136  
QY 3725 ATGAGAGCTTCTGGCTCTGAAGAGCTGACAGCTTGGCTGTGGCGCTCGGCCAAGAC 3784  
Db 135 ATGAGAGCTTCTGGCTCTGAAGAGCTGACAGCTTGGCTGTGGCGCTCGGCCAAGAC 76  
QY 3785 TTCAACCGGCTCAAGAAAGATGAGCTTCCAGAGCTGAGTACCCCTGACCTGGGG 3844  
Db 75 TTCAACCGGCTCAAGAAAGATGAGCTTCCAGAGCTGAGTACCCCTGACCTGGGG 16  
QY 3845 GCTCATGGCTTCTGA 3859  
Db 15 GCTCATGGCTTCTGA 1

RESULT 4

US-09-864-761-6462/c

;; Sequence 6462, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
;; FILE REFERENCE: Aemica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 6462  
;; LENGTH: 492  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC005849.1  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5  
US-09-864-761-6462

Query Match 9.7%; Score 492; DB 10; Length 492;  
Best Local Similarity 100.0%; Pred. No. 4.1e-122;  
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3519 TTCTGTCTTACTTGGTGGCTCAACGCTCAGGCTGACCTGCTGAGCTGCGCCAGCC 3578  
Db 492 TTCTGTCTTACTTGGTGGCTCAACGCTCAGGCTGACCTGCTGAGCTGCGCCAGCC 433

QY 3579 TGGCCACATTTGACACAGCTCCAGGGCTGCTGGGAGCATTTGGGGGTTCATGGCAG 3638  
Db 432 TGGCCACATTTGACACAGCTCCAGGGCTGCTGGGAGCATTTGGGGGTTCATGGCAG 373  
QY 3639 CTCTGGGCTACCCACTGCGCCCGCCAGCCGCTGCTGGGACTGAACCCACTTTGAGCTCTCTGGCC 3698  
Db 372 CTCTGGGCTACCCACTGCGCCCGCCAGCCGCTGCTGGGACTGAACCCACTTTGAGCTCTCTGGCC 313  
QY 3699 CTGCCCACAGTGAATTCCTCCAGAGATGACGACTTCTGGCTGCTGAAGAGAGCTGCAGA 3758  
Db 312 CTGCCCACAGTGAATTCCTCCAGAGATGACGACTTCTGGCTGCTGAAGAGAGCTGCAGA 253  
QY 3759 CTGGCTGTGGGCTGCGCCCGCCAGGACTTCAACCGGCTCAAGAGAGATGACGCTCCAG 3818  
Db 252 CCGTGTGTGGGCTGCGCCCGCCAGGACTTCAACCGGCTCAAGAGAGATGACGCTCCAG 193  
QY 3819 CAGCTGAGTCACTGACCTGCGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCT 3878  
Db 192 CAGCTGAGTCACTGACCTGCGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCT 133  
QY 3879 CGCTCCCTTCAAACTGCTCCACTTTGTGAGAGCCGCTGTATGCGCAACACTG 3938  
Db 132 CGCTCCCTTCAAACTGCTCCACTTTGTGAGAGCCGCTGTATGCGCAACACTG 73  
QY 3939 TTGAGCAGAGACAGAGCTGTGAGCTCTGCGCTTCTGAGCCGCTGGGCGGTG 3998  
Db 72 TTGAGCAGAGACAGAGCTGTGAGCTCTGCGCTTCTGAGCCGCTGGGCGGTG 13  
QY 3999 ATCGATGACCC 4010  
Db 12 ATCGATGACCC 1

RESULT 5  
US-09-867-701-4667  
; Sequence 4667, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Agiate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4667  
; LENGTH: 465  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-867-701-4667  
Query Match 8.6%; Score 437.2; DB 10; Length 465;  
Best Local Similarity 98.9%; Pred. No. 2.1e-107;  
Matches 461; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
QY 133 CAGGTTGAAACCCAACTAGCCCTGCTTTCATTAACATGACAAGAGCGCCCATCTGA 192  
Db 1 CAAGCTGAAACCCAACTAGCCCTGCTTTCATTAACATGACAAGAGCGC-CCCCATCTGA 59  
QY 193 TACCTTAACCCAGACAGGCTTCAACTCACTCCCTTCTGCTGCGCCAGACCTCACC 252  
Db 60 TACCTTAACCCAGACAGGCTTCAACTCACTCCCTTCTGCTGCGCCAGACCTCACC 119  
QY 253 CATCTTGTGAGCTCAAACTCAACCGCACTTAATCAACCAATCCCAAGTCTAACTA 311  
Db 120 CATCTTGTGAGCTCAAACTCAACCGCACTTAATCAACCAATCCCAAGTCTAACTA 179  
QY 312 ATCTGAACTTTAAAGTAAACCCAGTCTTAACTCACTAGCCCAATGCAATATATAT 371

Db 180 ATCTGAAACTTTTAAAGTAACCCAGTCTTTAAACCTTAACCTAGGCCAATGCAATTATAT 239  
QY 372 CTACCCCTAGCAACACCTTAACCTGCTTTGCGAGTCCAAAGTGTCCACTGAATCTCACT 431  
Db 240 CTACCCCTAGCAACACCTTAACCTGCTTTGCGAGTCCAAAGTGTCCACTGAATCTCACT 299  
QY 432 TGGTCTCTCACTGAAATCCCAAGAAAGATATTTCCCACTGCGCCACATCCCTCTTACA 491  
Db 300 TGGTCTCTCACTGAAATCCCAAGAAAGATATTTCCCACTGCGCCACATCCCTCTTACA 359  
QY 492 GCACCCAACTGCGCTTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 551  
Db 360 GCACCCAACTGCGCTTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419  
QY 552 TCACCCAACTGCGCTTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597  
Db 420 TCACCCAACTGCGCTTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 465  
RESULT 6  
US-09-931-704-4  
; Sequence 4, Application US/09931704  
; Patent No. US20020041873A1  
; GENERAL INFORMATION:  
; APPLICANT: Senaldi, Giorgio  
; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Usin  
; FILE REFERENCE: A-695  
; CURRENT APPLICATION NUMBER: US/09/931,704  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: US 60/226,436  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 819  
; TYPE: DNA  
; ORGANISM: Murine  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (95)..(769)  
; OTHER INFORMATION:  
; NAME/KEY: mat\_peptide  
; LOCATION: (176)..()  
; OTHER INFORMATION:  
; NAME/KEY: mat\_peptide  
; LOCATION: (176)..(769)  
; OTHER INFORMATION:  
; NAME/KEY: sig\_peptide  
; LOCATION: (95)..(175)  
; OTHER INFORMATION:  
US-09-931-704-4

Query Match 8.4%; Score 429.6; DB 10; Length 819;  
Best Local Similarity 90.4%; Pred. No. 3.3e-105;  
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 3363 AGCTGAACCTAGCTGGGCCCCCTTTTCAACGAGCAGACTTCAACCTCCCGCTGGGG 3422  
Db 276 ACCTGAACCTAGCTGGGCCCCCTTTTCAACGAGCAGCTGACTTCAATCTCTCGACTGGGG 335  
QY 3423 CAGAGACTTGGCCAGGCGCCACTTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAAAC 3482  
Db 336 CAGAAACTCTGCCCCAGGCGCCCGCTCACTTGGAGTGTGGGAGAGCTCAATGACAGGC 395  
QY 3483 TGGGGCTGACCCAGAACTACGAGGCTTACAGCCACTTCTGTGTGTACTTGGTGGCCTCA 3542  
Db 396 TGGGGCTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTGTACTTGGTGGCCTCA 455  
QY 3543 ACCGTGAGGTGCGACTGTGCTGAGCTGCGCGCAGCCTGCGCCACTTCTGACACGCTCC 3602  
Db 456 ACCGTGAGGTGCGACTGTGCTGAGCTGCGCGCAGCCTGCGCCACTTCTGACACGCTCC 515



```
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-5382

Query Match      4.9%; Score 250.6; DB 10; Length 269;
Best Local Similarity 98.1%; Pred. No. 1.8e-57;
Matches 264; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY   133 CAGGTTGAATACCAAACTAGCCCTGCTCTTCATAAATGACAAGCAGCGGCCCATCTGA 192
    |||
Db    1 CAAGCTGAAAACCCAAACTAGCCCTGCTCTTCATAAATGACAAGCAGCGGCCCATCTGA 60

QY   193 TACCTAAACCGAACCAAGTCAAGCCCTCCAACCTCACCCCTCTGCCTGGCCAGACCTCACCA 252
    |||
Db    61 TACCTAAACCGAACCAAGTCAAGCCCTCCAACCTCACCCCTCTGCCTGGCCAGACCTCACCA 120

QY   253 CATCCTTG-TGGACTCAAACCTCAACCGCAGCTAAATCAACAATCCCAAGTCTTAAACTA 311
    |||
Db    121 CATCCTTGCTGGACTCGAGACCTCAACCGCAGCTAAATCAACAATCCCAAGTCTTAAACTA 180

QY   312 ATCTGAAACTTTTTAAAGTAACCAGTCTCTTAAACCTTAAACCTTAGCCCAATGCAATTAT 371
    |||
Db    181 ATCTGAAACTTTTTAAAGTAACCAGTCTCTTAAACCTTAAACCTTAGCCCAATGCAATTAT 240

QY   372 CTACCCTAGCCAAACCCCTTAACCTGCGCTTTG 400
    |||
Db    241 CTACCCTAGCCAAACCCCTTAACCTGCGCTTTG 269

RESULT 10
US-09-867-701-2639
; Sequence 2639, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2639
; LENGTH: 231
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-2639
```

	Query Match	3.8%;	Score 192.8;	DB 10;	Length 231;
	Best Local Similarity	98.2%;	Prod. No. 5.5e-42;		
	Matches 216;	Conservative 2;	Mismatches 2;	Indels 2;	Gaps 2;
Qy	132	GCAGGTGGAACCCAAACTAGCCTGCTTCTTCATTAACATGACAGCAGCGCCCATCTG	191		
Db	13	GCAGCTGAAACCCAACTAGCCTGCTTCTTCATTAACATGACAGCAGC-CCCATCTG	71		
Qy	192	ATACCTAAACCGACCAAGTCAAGCCCTTCCAACTCACCTTGCCTGCCAGACCTCAC	251		
Db	72	ATACCTAAACCGACCAAGTCAAGCCCTTCCAACTCACCTTGCCTGCCAGACCTCAC	131		
Qy	252	ACATCCTTG-TGGACTCAAACTCAACCGCCTAAATCAACCAAAATCCCAAGTCTAAACT	310		
Db	132	ACATCCTTGCTGGACTCAAACTCAACCGCCTAAATCAACCAAAATCCCAAGTCTAAACT	191		
Qy	311	AATCTGAAACTTTTAAAGTAACCCAGTCTCTTAAACCTAAC	350		
Db	192	ATCTGAAACTTTTAAAGTAACCCAGTCTCTTAAACCTAAC	231		

RESULT 11  
US-09-765-231A-75

```

; Sequence 75, Application US/09765231A
; Patent No. US20020119452A1
; GENERAL INFORMATION:
; APPLICANT: Searle/Monsanto
; APPLICANT: Phippard, Deborah
; APPLICANT: Vasanthakamur, Geetha
; APPLICANT: Dotson, Stanton
; APPLICANT: Ma, Xiao-Jun
; TITLE OF INVENTION: Osteoarthritis tissue-derived nucleic acids, polypeptides,
; and cells
; TITLE OF INVENTION: vectors, and cells
; FILE REFERENCE: SO-3221 PR
; CURRENT APPLICATION NUMBER: US/09/765,231A
; CURRENT FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 82
; SEQ ID NO 75
; LENGTH: 283
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-765-231A-75

```

	Query Match	3.2%;	Score 165.2;	DB 10;	Length 283;
	Best Local Similarity	86.9%;	Prod. No. 1.6e-34;		
	Matches 193;	Conservative	0;	Mismatches 28;	Indels 1; Gaps
QY	1328	CCCCGGCCGGCTCATGTCTTCGTCCTGCATCTGCCACACAGGGACT-CGTGGGGGATGTT	1386		
DB	30	CCCCGCCGCCGCCCGCCACGCCCATGGACCTCCGAGCAGGGGACTCGGTGGGGATGTT	89		
QY	1387	AGCGTGCCCTGTGCAGGTGCTCTGGCAGCTCCCTCGTGCAGTGCAGCTCTCAATGCACAGG	1446		
DB	90	AGCGTGCCCTGTGCACGGTGCTCTGGCACCTCCCTGTCAGTGCACAGCTCTCAATGCACAGG	149		
QY	1447	GSACCCAGGCGCTGCGCCCTCCATCCAGAAAACTATGACCTCACCCGCTACCTGGAGCA	1506		
DB	150	GSACCCAGGCGCTGCGCCCTCCATCCAGAAAACTATGACCTCACCCGCTACCTGGAGCA	209		
QY	1507	CCAACTCCGACGTTGGCTGGGACCTATGTAGTATCCACGG	1548		
DB	210	CCAACTCCGACGTTGGCTGGGACCTATCTGAACCTACCTGGG	251		

RESULT 12  
US-09-904-968A-1  
; Sequence 1, Application US/09904968A  
; Publication No. US2003008288A1  
; GENERAL INFORMATION:  
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
; APPLICANT: GERMINO, Gregory  
; APPLICANT: WATNICK, Terry  
; APPLICANT: PHADDEKITCHAROEN, Bunyong  
; TITLE OF INVENTION: DETECTION AND TREATMENT OF POLYCYSTIC KIDNEY DISEASE  
; FILE REFERENCE: JHU1680-2  
; CURRENT APPLICATION NUMBER: US/09/904,968A  
; CURRENT FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/283,691  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/218,261  
; PRIOR FILING DATE: 2000-07-13  
; NUMBER OF SEQ ID NOS: 113  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 53522  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-904-968A-1

Query Match 1.1%; Score 57.6; DB 9; Length 53522;  
Best Local Similarity 45.4%; Pred. No. 0.00046;  
Matches 367. Conservative 0; Mismatches 429; Indels 12;

QY 2319 CCTCCCCCTGGGCGCCCTTGCCATCCAGGCCCTCCTCCCCTGTCTTCCCTTCCTTTCCAG 2378  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 34289 CTTCTCCCTCCCTAGCCCTTTCCCTTCTTCCCCCCCAGGCCCTTCCCTCCTCCCTC 34348

Qy	2379	TTATACATCTCCCTCATCCCTTTCCTCGGGGCCAGCGCTCCCCGAGGGTGTAAGAAGG	2438
Db	34349	CCCTAGCGCTTCCCCTCCTCCCTCCTCCCTAAGCCCTTCCCTCCTGC--CCTCCCTAGAC	34406
Qy	2439	GCTGTGGCTCTTCCCTATACCAATGCTGCTTTCATAGACCTTCTCTCTGTCCTACTCATG	2498
Db	34407	CTTCCCTCACTCTCTCCCGTAGCGCCCTCACACTGTCGCCAGCCCTCCCTCCCTCA	34466
Qy	2499	AGACTGCCTCAATTCTTCTTCTTGCAACCCTGTCTCTTATCAGTGAACCTCTTCTTCGG	2558
Db	34467	GCCCTCCCTCCCTTCTCTCCCTCCTCCCTCCTCCTCCCTCCTCCTCCTCTCTCTC	34526
Qy	2559	AGTGTAGTAGTACCGCTCTCTCCCAAGCCCTCAGCTGTGTGGCCCTGTGGTGTACG	2618
Db	34527	CCCTCCCTCCTCCTCCCTTCTCCCTCTCTCCTCCCT-----CCCTCCTGTCTCCCTC	34581
Qy	2619	GGCAAATGGGGCTGTGTGTCCAATAGGGGCATCTCATCTCTCTTGTCTCTTGTGCAGA	2678
Db	34582	TCCCTGCCCTCTCCCTCTCCTCCCTCCTCCCTCCTCCTCCCTCCTCCTCCTCCTCCT	34641
Qy	2679	AACCTTTGCTTCACTCCACATGCCCTCTCTAGTTCGCCAGCCCTTTTCTCTCCTGGCTTT	2738
Db	34642	CCTCCCTCCTCTCCTCCTCCCTCCTCCTCCTCCTCCCTCCTCCCTCCTCCTCCTCCT	34701
Qy	2739	CCCTGCCAAATTTCTCAAGGAGTGTCTACACCTCTGC--CTCCACTTCTCTCTCCACC	2796
Db	34702	CCCCCTCCCTTCTCCCTCCCTCCCTCCCTCCCTCCTCCTCCCTCCTCCTCCTCCCTC	34761
Qy	2797	CACTCACCTTTTAACCCCTCGAATCTGGCTTCAGGCGCCAGCAATGGTCTCTCCAG	2856
Db	34762	CTCCATCCCTTCCTCCCGTTCCCAATCTCTCCCTCCCTCCCTTCCATTTCTCCTCTCC	34821
Qy	2857	GTCTGTAGGCACCTCTTGCCAAAGCCGACAGTGTTTTGAAGGCTCATCTCCTTGCTGT	2916
Db	34822	CCCTGCCCTCCTCTCCTCTCA---CCTCCCTCTCTCGCTCCTTCTTCTCTCCTCCTCC	34878
Qy	2917	CTGTTTTGACGCCACACTGTGAGCGCTGTGCTCTTCGAACCTCTCTTCTTGGTCTC	2976
Db	34879	CTTTCTCTCTCCCTCCCTTCTCCCTTCTCCTCTTCTCCCTTCTCCTCTCTTTTTCAT	34938
Qy	2977	TGCACCTCTCTGGGCACCTTCTACCTCTCAGCTCTCTCCAGGCTCTCTTCTCTCTCTGT	3036
Db	34939	CCTTCCCTCTCTCCCTCTCTTCTCTCTCTTTTCTCCTCTTCTTCTCCCTCCCTCCTCT	34998
Qy	3037	CCTGCCCCACAGGGGGACATCTCCCAAGTTTGGCCACCCAGCCAAATCAGACGTCTCT	3096
Db	34999	CCTCCTCCCATTCCTCCCTCCTCCCTCCCTCCCATTCCTCCCTCTCTCTCTCTCTCC	35058
Qy	3097	CCTGAGGCTCTGTGCGTCTCCTCCTCC	3124
Db	35059	CATTACCCCTCTCTCTCCCTCCTCTCC	35086

```

RESULT 13
US-09-804-682-29
; Sequence 29, Application US/09804682
; Patent No. US20020106765A1
; GENERAL INFORMATION:
; APPLICANT: Kinders, Robert
; APPLICANT: Corey, Michael J.
; TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
; TITLE OF INVENTION: MODULATING THE SAME
; FILE REFERENCE: 130001.406
; CURRENT APPLICATION NUMBER: US/09/804,682
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1064
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

:
:
: NAME/KEY: misc:feature
: LOCATION: 5, 6, 16, 21, 24, 25, 33, 39, 72, 110, 209, 214, 231, 232,
: LOCATION: 235, 237, 238, 244, 245, 246, 256, 282, 292, 297, 306, 319,
: LOCATION: 321, 323, 330, 334, 340, 349, 354, 355, 363, 372, 376, 378,
: LOCATION: 397, 405, 432, 437, 454, 455, 457, 458, 459, 468, 470
:
: OTHER INFORMATION: n = A,T,C or G
:
: NAME/KEY: misc:feature
: LOCATION: 485, 487, 488, 494, 496, 499, 511, 524, 527, 552, 557, 562,
: LOCATION: 583, 600, 611, 613, 623, 624, 652, 654, 674, 681, 687, 691,
: LOCATION: 694, 701, 713, 716, 720, 721, 725, 731, 734, 735, 739, 743,
: LOCATION: 744, 781, 782, 785, 789, 799, 803, 821, 823, 847, 852
:
: OTHER INFORMATION: n = A,T,C or G
:
: NAME/KEY: misc:feature
: LOCATION: 858, 878, 884, 886, 896, 897, 901, 917, 926, 932, 939, 948,
: LOCATION: 957, 961, 965, 981, 994, 993, 1001, 1002, 1005, 1011, 1018,
: LOCATION: 1043, 1047, 1049, 1051, 1054, 1056
:
: OTHER INFORMATION: n = A,T,C or G
:
: US-09-804-682-29

```

	Query Match	1.1%	Score 54.4;	DB 10;	Length 1064;
	Best local Similarity	39.0%;	Pred. No. 0.00023;		
	Matches 252;	Conservative	0;	Mismatches 393;	Indels 1;
	1;	Gaps	1;		
Qy	2432	GGAAAGGCGCTGCGCCCTCTCCCTATACCATGCTGTCTTCCATAGCGCTTCTCTCGTGCCT	2491		
Db	266	GGTGATGAGGACACCTTCTTCCCCACGNTCCGCTCACTTCTTCCCGCTCCGNGTNCNC	325		
Qy	2492	ACTCATGAGACTGCGCTCCATTTCTTCTTCTTGTGCAACCGCTGTCTCTATCAGCTGAACCGTT	2551		
Db	326	CCCTNATNCNCCCTTNCNCCCTCCTTCCNCCCTCAANTCTTCCCNCTCTNNACTCCCC	385		
Qy	2552	CTTTGCGAGTGTAGTGAGTACCGGTCTCTTCCCGAGCCCTCAGCTGGTGGCGCTGGGTG	2611		
Db	386	CCCCCTTCCGNCNCCCTCCNAGCCCTCTCCGCCCCCTTCCGCTTCCGNCNCCGCCCC	445		
Qy	2612	TGTCAGGCGCAATGGGGCTCTGGTTCCAAATGGGCGACTCTCATCTCTCTTGTGCTT	2671		
Db	446	CTTC- CCGCCNCCNCCCTCNCNTCCGCCCCCCCCCNCNCCCTCNCNCTCCTCTC	504		
Qy	2672	GTGCAGAAACCTTTGTCTCACTCCAGTGGCTCTCTAGTTCCGAGCCCTTTTCTCTCC	2731		
Db	505	CCCCCNCCCTCCCCCGCNCNCCGCCCCCTCTCTA CCGCTCNCNCCGNCNCC	564		
Qy	2732	TGGCTTTCCCTGCGAAATTTCTCCAAGGAGTGTGTACAGCTCTGCTGCGACTTCCTCT	2791		
Db	565	CAACCCGCCCCCGCTCCTCTCCCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC	624		
Qy	2792	CCACGCACTCACTTCTTAAACCCCTGTGAATCTGGCTTCCAGGCCCCAGCAATGTTCTCT	2851		
Db	625	TCCGCTCTCCCTCCGCCCCGCCCCGNCNCCGCTCTCTCCGCCCCGNCNCCCTCC	684		
Qy	2852	CCAAGGTGCTCAGGCACTCTTGGCAAGCCGACAGTGTTTGAAGGCTCATCTCTCTT	2911		
Db	685	CCNCTTNCNCCGCCCCCTCAGCTCNCNCCNCTCNCCTCNCNCCNCCNCCN	744		
Qy	2912	GCTGTCTGTTTGGAGCGCACTGCTGAGGCTGCTGCTTCTCGAACTCTCTTCTGTTG	2971		
Db	745	CTTCCGCGCGTACCCGCGCTTCCCTTCCGCCCCNCCNCCNCTCCCCCCCCNCCNC	804		
Qy	2972	GTCTCTGCATCTCTCTGGGCGACCTTCTAGCTTCTCAGCTCTCTCAGGCTCTCTTCCCTC	3031		
Db	805	CCCCGCTCTCCGCCCCNCCGCTCTCTCTCTCCCTTCCNCCCTCCNCCCTCCCTCCTC	864		
Qy	3032	TCTGTCTGCCCCCAGCGGGGCACTCTCCCAAGGTTTGGCCACC	3077		
Db	865	CTCACCTCCGCGCTCTCTTCCNCCCTCCGCGNCCCTTCTCTCTCTCTCTCTCTCTCT	910		

RESULT 14  
US-09-864-761-10689  
; Sequence 10689, Application US/09864761  
; Patent No. US20020048763A1

```

; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US 09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10689
; LENGTH: 496
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005296.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.91
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; US-09-864-761-10689

Query Match 1.0%; Score 53; DB 10; Length 496;
Best Local Similarity 49.0%; Pred. No. 0.00033;
Matches 171; Conservative 0; Mismatches 175; Indels 3; Gaps 1;

QY 2796 CCACCTCACTTCTTAACCCCTGCAATCGCTTCCAGGCCCAAGCAATGGTTCTCTCCAA 2855
Db 105 CCACCTTCTCTTCTTCTCTCTCTCTCTTGGTTTCTTCTCTCTCTCTCTCTCTCTCT 164

QY 2856 GGTCGTAGGCACCTCTCTTGGCAAGCCGACAGTGTTTGAAGGCTCATCTCTCTGCTG 2915
Db 165 CTTCCTTCTCTCCCTTACTCTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 224

```

```

Qy 2916 TCTGTTTTCGAGCCACACTGCTGAGCGTGCCTTCTTCGAACTCCTCTTC---CTTGG 2972
Db 225 TCTTTTCTCCTTCTCCTTTTCTTTTCCCTTCTCCTCTCCTCTCCTCTCCTCTCTCC 284
Qy 2973 TCTCTGCACTCTCTCGGGCCACCTTCTACCTCTCAGCTCCTCCAGGCTCCCTCTCCTCT 3032
Db 285 TCTTTTCTCCTCCCTCTCCTCTCTCCCTCCCTCTACTCCTTCTCTCTCCTCTTCCCTC 344
Qy 3033 CTGTCCTCTGCCCCACAGCGGGGCACTCTCCCAAGTTTGCCACCCAGCAATCAGCACGT 3092
Db 345 CCTCCTCTCTTCTCCTCCTCCCTCTCTTCTCCCTCCTCCTTCTCTCTCTCTCTCCTCT 404
Qy 3093 CTTTCTCTGAGCGTCTTGTCGCTCTCCTCTCCTCTCCTCTTCTTCTTCTACGCGCTC 3141
Db 405 CTTTCTTCTCTTCTCCTCTTTTCTCCTCTCTTCTCCCTCTTCTCCCTCCCGCGCTC 453

RESULT 15
US-09-804-682-33
; Sequence 33, Application US/09804682
; Patent No. US20020106765A1
; GENERAL INFORMATION:
; APPLICANT: Kinders, Robert
; APPLICANT: Corey, Michael J.
; TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
; TITLE OF INVENTION: MODULATING THE SAME
; FILE REFERENCE: 130001.406
; CURRENT APPLICATION NUMBER: US/09/804,682
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14, 21, 27, 33, 35, 42, 72, 101, 103, 207, 208, 221, 223,
; LOCATION: 227, 235, 240, 242, 243, 247, 248, 259, 263, 269, 273, 278,
; LOCATION: 296, 321, 322, 324, 330, 332, 335, 336, 340, 367, 371, 385,
; LOCATION: 390, 393, 399, 401, 407, 415, 421, 447, 454, 471, 475
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 479, 494, 501, 508, 511, 513, 525, 539, 540, 579, 596, 605,
; LOCATION: 623, 624, 633, 653, 654, 656, 658, 664, 670, 679, 711, 713,
; LOCATION: 721, 723, 735, 743, 747, 748, 754, 776, 778, 779, 780, 783,
; LOCATION: 798, 802, 808, 810, 813, 814, 820, 822, 824, 825, 838
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 842, 847, 888, 900, 909, 910, 916, 926, 927, 943, 948, 962,
; LOCATION: 973, 1002, 1005, 1028, 1029, 1034, 1057, 1065
; OTHER INFORMATION: n = A,T,C or G
US-09-804-682-33

```

	Query Match	1.0%	Score 53;	DB 10;	Length 1065;
	Best Local Similarity	39.1%;	Pred. No. 0.00055;		
	Matches 318;	Conservative 0;	Mismatches 488;	Indels 8;	Gaps 2;
QY	2332	CCTTGCCCTATCCGAGGCTCTCCCTTGCTCTCTCCCTCTCTTCCAGTTATACATCTCCC	2391		
DB	255	CGGNCCTTNTCCACNCCNCACNCTAAATTTCCACTTTCCNCAAGCACTCATCTCCCC	314		
QY	2392	TCATGCCCTTTCCCTGGGCCCCGAGCGCTCCCCCGAGGGTTGGAAAGGCGCTGCGCCTCTT	2451		
DB	315	CCCCCNCNTNCCCGTNCNCCNCCNCCGCCACCTCTCTCCCCCCCCCNCTCNCCC	374		
QY	2452	CCCTATACCATGCTGTCTTTCATAGCTTCTCTCGTGCTTACTCATGAGACTGCGCTCCAT	2511		
DB	375	CCCTCTCTTCCNCCNCCNCCNCTCTNCTNCTCTNCCCTCTTNTTCAANCCCCCCCCCCC	432		
QY	2512	TTCTTCCTTCTGCAACCCCTGCTCTCATCAGCTGAACCCCTTCTTTTCGGAGTGTTAGTGAGT	2571		







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 163.925 Seconds  
(without alignments)  
9516.918 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 5087

Sequence: 1 aacctgcgagtgccctggc.....ccttgttaagcttctcca 5087

Scoring table: IDENTITY NUC

Gapop 10'0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5087	100.0	5087	1	US-08-792-019B-3
2	5087	100.0	5087	3	US-08-988-819-3
3	5087	100.0	5087	3	US-08-988-819-3
4	1463.4	28.8	1710	3	US-09-016-534-3
5	523.4	10.3	797	1	US-08-792-019B-1
6	523.4	10.3	797	3	US-08-988-819-1
7	523.4	10.3	797	3	US-08-988-819-1
8	429.6	8.4	819	1	US-08-792-019B-4
9	429.6	8.4	819	3	US-08-988-819-4
10	429.6	8.4	819	3	US-08-988-819-4
11	324.6	6.4	396	3	US-09-016-534-4
12	99	1.9	7218	1	US-08-106-182-7
13	60.2	1.2	1926	4	US-08-232-463-14
14	60.2	1.2	1926	4	US-08-232-463-14
15	57.6	1.1	53526	3	US-09-130-114-2
16	57.6	1.1	53577	3	US-08-658-136-2
17	52.4	1.0	51259	3	US-08-658-136-1
18	52.2	1.0	1166	4	US-08-781-891-209
19	49.6	1.0	16442	3	US-09-072-596-323
20	48.8	1.0	1926	4	US-08-781-891-208
21	48.8	1.0	1926	4	US-09-249-585A-2
22	48.8	1.0	2580	3	US-09-050-863-2
23	48.8	1.0	2580	4	US-09-359-081-2
24	48.8	1.0	5452	2	US-09-130-114-1
25	48.8	1.0	7218	1	US-08-232-463-14
26	48.8	1.0	9600	4	US-08-910-647-1
27	48.8	1.0	9600	4	US-08-910-647-1
28	48.8	1.0	10596	1	US-07-884-811-15

C 28	48.8	1.0	10596	1	US-07-885-971-15	Sequence 15, Appl
C 29	48.8	1.0	10596	1	US-08-087-783A-15	Sequence 15, Appl
C 30	48.8	1.0	10596	1	US-08-194-088B-15	Sequence 15, Appl
C 31	48.8	1.0	10596	2	US-08-194-087-15	Sequence 15, Appl
C 32	48.8	1.0	10596	5	PCT-US93-04648-15	Sequence 15, Appl
C 33	48.6	1.0	289	4	US-09-007-005-17	Sequence 17, Appl
C 34	48.6	1.0	289	4	US-09-244-796-17	Sequence 17, Appl
C 35	48	0.9	1866	3	US-09-173-581-13	Sequence 13, Appl
C 36	48	0.9	1866	4	US-09-420-915-13	Sequence 13, Appl
C 37	45.2	0.9	950	4	US-08-991-789A-6	Sequence 6, Appl
C 38	45.2	0.9	950	4	US-09-062-451-6	Sequence 6, Appl
C 39	45.2	0.9	950	4	US-09-598-326-6	Sequence 6, Appl
C 40	44.8	0.9	68750	3	US-09-335-409-1	Sequence 1, Appl
C 41	44.8	0.9	68750	4	US-09-568-102-1	Sequence 1, Appl
C 42	44.8	0.9	68750	4	US-09-567-969-1	Sequence 1, Appl
C 43	44.8	0.9	68750	4	US-09-568-480-1	Sequence 1, Appl
C 44	44.8	0.9	68750	4	US-09-568-486-1	Sequence 1, Appl
C 45	44.8	0.9	68750	4	US-09-568-472-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-08-792-019B-3  
; Sequence 3, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUROTROPIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/792,019B  
; FILING DATE: 03-FEB-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5087 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 137..138  
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED  
; OTHER INFORMATION: REGION OF >1KB"  
US-08-792-019B-3

Query Match 100.0%; Score 5087; DB 1; Length 5087;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5087; Conservative 0; Mismatches 0; Gaps 0;  
Qy 1 AACCTGCGAGTGGCGCTGGCGGATGGGATTTAAAGCTTCGCGGAGCCCGGCTGCC 60  
Db 1 AACCTGCGAGTGGCGCTGGCGGATGGGATTTAAAGCTTCGCGGAGCCCGGCTGCC 60

QY 61 CTCCTCACTCCGAGCCTCCGGAGAGAGCGCACCCCGCGCGCCAGCCCGCCCA 120  
Db 61 CTCCTCACTCCGAGCCTCCGGAGAGAGCGCACCCCGCGCGCCAGCCCGCCCA 120  
QY 121 TGGAGCTCCGAGCAGGTTGAAAACCCAAACTAGCCCTGCTCTTTTATAAATGAAAGCAG 180  
Db 121 TGGAGCTCCGAGCAGGTTGAAAACCCAAACTAGCCCTGCTCTTTTATAAATGAAAGCAG 180  
QY 181 CGCCCATCTGATACCTAAACGACCAAGTCACAGCGCTCCCAACTCACCTCTGCGCTGCC 240  
Db 181 CGCCCATCTGATACCTAAACGACCAAGTCACAGCGCTCCCAACTCACCTCTGCGCTGCC 240  
QY 241 CAGAGCTCACCAATCTTTGTGAGCTCAAACTCAACCGCACTAAATCAACCAAAATCCCA 300  
Db 241 CAGAGCTCACCAATCTTTGTGAGCTCAAACTCAACCGCACTAAATCAACCAAAATCCCA 300  
QY 301 AGTCTAACTAATCTGAAACTTTTAAAGTAAACCCAGTCTTTAAACCTTAACCTAGCCCAAT 360  
Db 301 AGTCTAACTAATCTGAAACTTTTAAAGTAAACCCAGTCTTTAAACCTTAACCTAGCCCAAT 360  
QY 361 GCGAATTATATCTACCTAGCCAAACCCCTAACTGCTTTGCGAGTCCAAAGTGTCCACTG 420  
Db 361 GCGAATTATATCTACCTAGCCAAACCCCTAACTGCTTTGCGAGTCCAAAGTGTCCACTG 420  
QY 421 AATCCTCACTTGGTCTCTACTGAAATCCAGAAAGCATATTTCCCACTGCCCACAT 480  
Db 421 AATCCTCACTTGGTCTCTACTGAAATCCAGAAAGCATATTTCCCACTGCCCACAT 480  
QY 481 CCTCTCTTACAGCACCAACCTTGGCTCTGGAATCTCTGGTATCTCTGGAGTGTCCAAACT 540  
Db 481 CCTCTCTTACAGCACCAACCTTGGCTCTGGAATCTCTGGTATCTCTGGAGTGTCCAAACT 540  
QY 541 GTGAGTGCATCAGCAACAGCCGACTCGTCAATGACCTCTCTCTCTCTCTCTCTCTCT 600  
Db 541 GTGAGTGCATCAGCAACAGCCGACTCGTCAATGACCTCTCTCTCTCTCTCTCTCTCT 600  
QY 601 CCACCTTGCAGGCTGATGAAAGGCTCATTTGAAGTCCAACTTTTCCCACTTAAACAC 660  
Db 601 CCACCTTGCAGGCTGATGAAAGGCTCATTTGAAGTCCAACTTTTCCCACTTAAACAC 660  
QY 661 AAGAACGGGTGAACTTCCACCTGCGCACCGTTCCTCTGAGAGTGAACATAAATCTCTCT 720  
Db 661 AAGAACGGGTGAACTTCCACCTGCGCACCGTTCCTCTGAGAGTGAACATAAATCTCTCT 720  
QY 721 CAATCTAACCCCACTTACATCTTCCCACTTCCCACTTCCCACTTCCCACTTCCCACT 780  
Db 721 CAATCTAACCCCACTTACATCTTCCCACTTCCCACTTCCCACTTCCCACTTCCCACT 780  
QY 781 ACTAAGCCCCATAGGAGCGAGCCGACCTAGTGTCTTAACTTATACCTTGTCTCTATGG 840  
Db 781 ACTAAGCCCCATAGGAGCGAGCCGACCTAGTGTCTTAACTTATACCTTGTCTCTATGG 840  
QY 841 GTGAGTCTGTCTTGGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
Db 841 GTGAGTCTGTCTTGGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
QY 901 AGCTGCGAGCTGTGATGTGTCTTCCCACTTCCCACTTCCCACTTCCCACTTCCCACT 960  
Db 901 AGCTGCGAGCTGTGATGTGTCTTCCCACTTCCCACTTCCCACTTCCCACTTCCCACT 960  
QY 961 GACTGGAAGACTTGGCAGGAGTAGGTTACAACTGGAAACAGGAGGTCGACCTGCGAGT 1020  
Db 961 GACTGGAAGACTTGGCAGGAGTAGGTTACAACTGGAAACAGGAGGTCGACCTGCGAGT 1020  
QY 1021 CCTTAGGCTTGGCGCGCGCT 1080  
Db 1021 CCTTAGGCTTGGCGCGCGCT 1080  
QY 1081 GCACATGCGAAGACTCTCTAGCTGACACAGATCCCACTTCTCAAGATCTACTGAT 1140  
Db 1081 GCACATGCGAAGACTCTCTAGCTGACACAGATCCCACTTCTCAAGATCTACTGAT 1140  
QY 1141 AGACACTCATGCGTGCAGAGTCTCTCATCTCAAAACATACATATGCTCTCTCTCTCTCT 1200

Db 1141 AGACACTCATGCGTGCAGAGTCTCTCATCTCAAAACATACATATGCTCTCTCTCTCTCT 1200  
QY 1201 GTCTTGCAGAGAGTCTTCCCTCTCTCATCTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 1260  
Db 1201 GTCTTGCAGAGAGTCTTCCCTCTCTCATCTCCCTCTCTCTCTCTCTCTCTCTCTCT 1260  
QY 1261 TCACCCCCCAACCCAGCCCAAGTGGGACACACCTGAGGGGCTGCCAGTCTCTCTCTCT 1320  
Db 1261 TCACCCCCCAACCCAGCCCAAGTGGGACACACCTGAGGGGCTGCCAGTCTCTCTCTCT 1320  
QY 1321 GTGTGGGCGCGCGCGCTCATGCTTCTGTCATCTCTGTCATCTCTGCCCCACAGGGGACTCTGGGG 1380  
Db 1321 GTGTGGGCGCGCGCGCTCATGCTTCTGTCATCTCTGTCATCTCTGCCCCACAGGGGACTCTGGGG 1380  
QY 1381 GATGTTAGCGTGTCTGTGACGCTCTCTGTCATCTCTGTCATCTCTGCCCCACAGGGGACTCTCAATCG 1440  
Db 1381 GATGTTAGCGTGTCTGTGACGCTCTCTGTCATCTCTGTCATCTCTGCCCCACAGGGGACTCTCAATCG 1440  
QY 1441 CACAGGGGACCCAGGGGCTGCGCTCTCATCCAGAAAACTTATGACCTCACCGCTACCT 1500  
Db 1441 CACAGGGGACCCAGGGGCTGCGCTCTCATCCAGAAAACTTATGACCTCACCGCTACCT 1500  
QY 1501 GGAGCAACAACTCCGAGCTTGGCTGGGACCTATGTGAGTATCCAGCTGAGGAATCTGGG 1560  
Db 1501 GGAGCAACAACTCCGAGCTTGGCTGGGACCTATGTGAGTATCCAGCTGAGGAATCTGGG 1560  
QY 1561 AGTTGGGAGGAGTGGGAGTGGGAAAGACAGTCTTAAACCGTGGAGGGTCTCTGTGATAA 1620  
Db 1561 AGTTGGGAGGAGTGGGAGTGGGAAAGACAGTCTTAAACCGTGGAGGGTCTCTGTGATAA 1620  
QY 1621 TGATGGGAGGAGGGGCTCTTTGGCTCCCAACAGTCCCTCTGTCTGTCTATCTCTCTG 1680  
Db 1621 TGATGGGAGGAGGGGCTCTTTGGCTCCCAACAGTCCCTCTGTCTGTCTATCTCTCTG 1680  
QY 1681 CCTTCTCTCTTGTGGGCGCGCGCTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1740  
Db 1681 CCTTCTCTCTTGTGGGCGCGCGCTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1740  
QY 1741 GGAGGCTCTCACCGCTTGGGCGCGCTTGGGCGCGCTTGGGCGCGCTTGGGCGCGCTTGGGCG 1800  
Db 1741 GGAGGCTCTCACCGCTTGGGCGCGCTTGGGCGCGCTTGGGCGCGCTTGGGCGCGCTTGGGCG 1800  
QY 1801 TCCCT 1860  
Db 1801 TCCCT 1860  
QY 1861 GGCGGGTGGGTGGGAGACAGGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1920  
Db 1861 GGCGGGTGGGTGGGAGACAGGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1920  
QY 1921 TGCCCCCAGACTGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1980  
Db 1921 TGCCCCCAGACTGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1980  
QY 1981 CCCATCTAGCTGGGCGCTCTCTAGGGGGTCTATGGGGGAGGGGATCTGTAGGGAACCCAGG 2040  
Db 1981 CCCATCTAGCTGGGCGCTCTCTAGGGGGTCTATGGGGGAGGGGATCTGTAGGGAACCCAGG 2040  
QY 2041 CAGTAGTGGCAGGGGTTTGGGTGGAGTGGAGTGTATGCTGTAGGATTTGGGGGTGG 2100  
Db 2041 CAGTAGTGGCAGGGGTTTGGGTGGAGTGGAGTGTATGCTGTAGGATTTGGGGGTGG 2100  
QY 2101 TCCAGAGTGTTCAGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2160  
Db 2101 TCCAGAGTGTTCAGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2160  
QY 2161 ATGGGGAACCGCGCGCT 2220  
Db 2161 ATGGGGAACCGCGCGCT 2220  
QY 2221 CCAGGGAAGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2280

Db 2221 CAGGGAAGAAAGGAAGAGGTTGGCGGGGAGCTGGCTCCAGCCCCAGGATACACCG 2280  
Qy 2281 AGGAAATAGTTTGTCTCTGTGTGTGACGCTGTGAACCTCCCTGGGCCCCCTTGCTA 2340  
Db 2281 AGGAAATAGTTTGTCTCTGTGTGTGACGCTGTGAACCTCCCTGGGCCCCCTTGCTA 2340  
Qy 2341 TCCAGGCTCTCCCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2400  
Db 2341 TCCAGGCTCTCCCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2400  
Qy 2401 TCCCTGGGCCCCAGCGCTCCCGAGGGTTGAAAGGGCTCTGCGCTCTCTCTCTCTCT 2460  
Db 2401 TCCCTGGGCCCCAGCGCTCCCGAGGGTTGAAAGGGCTCTGCGCTCTCTCTCTCTCT 2460  
Qy 2461 ATGCTGTCTTCCATAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520  
Db 2461 ATGCTGTCTTCCATAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520  
Qy 2521 CTGCAACCTGCT 2580  
Db 2521 CTGCAACCTGCT 2580  
Qy 2581 TCCCAAGCCCCCTCAGCTGTGGGCTTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2640  
Db 2581 TCCCAAGCCCCCTCAGCTGTGGGCTTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2640  
Qy 2641 ATGGGCACT 2700  
Db 2641 ATGGGCACT 2700  
Qy 2701 CCTCTCTAGTTCCCGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760  
Db 2701 CCTCTCTAGTTCCCGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760  
Qy 2761 GTGGTCTACACCTCTGCT 2820  
Db 2761 GTGGTCTACACCTCTGCT 2820  
Qy 2821 TCTGGCTTCCAGGCCCCAGCAATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880  
Db 2821 TCTGGCTTCCAGGCCCCAGCAATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880  
Qy 2881 CCCGACAGTTTGAAGCTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940  
Db 2881 CCCGACAGTTTGAAGCTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940  
Qy 2941 CGTCTGCT 3000  
Db 2941 CGTCTGCT 3000  
Qy 3001 CCTCTCCAGCTCTCCAGGCT 3060  
Db 3001 CCTCTCCAGCTCTCCAGGCT 3060  
Qy 3061 CCAAGTTTGGCCACCCAGCCCAATCAGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3120  
Db 3061 CCAAGTTTGGCCACCCAGCCCAATCAGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3120  
Qy 3121 CTCCT 3180  
Db 3121 CTCCT 3180  
Qy 3181 ACCTGCATACAAATGATATCTTATGGAATACTCAGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3240  
Db 3181 ACCTGCATACAAATGATATCTTATGGAATACTCAGGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3240  
Qy 3241 TAGCATGAGACAGGGCCAGT 3300  
Db 3241 TAGCATGAGACAGGGCCAGT 3300  
Qy 3301 TCTCTTGGTGTGAGCCAGGGCTCTGCT 3360  
Db 3301 TCTCTTGGTGTGAGCCAGGGCTCTGCT 3360

Qy 3361 ACAGCTGAACCTACTCGGCCCCCTCTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGG 3420  
Db 3361 ACAGCTGAACCTACTCGGCCCCCTCTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGG 3420  
Qy 3421 GGCAAGAGCTCTGCCCCAGGGCCACTGTGACTTGGAGGTGTGGGAGCCCTCAATGACAA 3480  
Db 3421 GGCAAGAGCTCTGCCCCAGGGCCACTGTGACTTGGAGGTGTGGGAGCCCTCAATGACAA 3480  
Qy 3481 ACTGCGGCTGACCCAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGGGTGGCCCT 3540  
Db 3481 ACTGCGGCTGACCCAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGGGTGGCCCT 3540  
Qy 3541 CAAACCTCAGGCTGCCACTGTGAGCTGCGCGCAGCCCTGCGGCTTCTGTGACAGCCCT 3600  
Db 3541 CAAACCTCAGGCTGCCACTGTGAGCTGCGCGCAGCCCTGCGGCTTCTGTGACAGCCCT 3600  
Qy 3601 CCAGGCTCTGCTGGGCGCACTTCTGGCTGCTGAAGGAGCTGCGGCTCTGGGCTACCCACTGCCCA 3660  
Db 3601 CCAGGCTCTGCTGGGCGCACTTCTGGCTGCTGAAGGAGCTGCGGCTCTGGGCTACCCACTGCCCA 3660  
Qy 3661 GCGCTGCTGGGCTGAGACCTTGGACTCTGCGCTCTGCGCTCTGCGCTCTGCGCTCTGCGCTCT 3720  
Db 3661 GCGCTGCTGGGCTGAGACCTTGGACTCTGCGCTCTGCGCTCTGCGCTCTGCGCTCTGCGCTCT 3720  
Qy 3721 GAAAGTGGACGACTTCTGGCTGCTGAAGGAGCTGCGGCTCTGCGCTCTGCGCTCTGCGCTCT 3780  
Db 3721 GAAAGTGGACGACTTCTGGCTGCTGAAGGAGCTGCGGCTCTGCGCTCTGCGCTCTGCGCTCT 3780  
Qy 3781 GGAATTCACCCGCTCAAGAAAGATGAGCTCCAGAGCTCCAGAGCTGAGTCAACCTGACCT 3840  
Db 3781 GGAATTCACCCGCTCAAGAAAGATGAGCTCCAGAGCTCCAGAGCTGAGTCAACCTGACCT 3840  
Qy 3841 GGGGCTCATGCTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCT 3900  
Db 3841 GGGGCTCATGCTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCT 3900  
Qy 3901 CCCACTTGTGAGAGCCAGCCCTGTATGCCAACACTGTTGAGCCAGGAGACAGAGCTG 3960  
Db 3901 CCCACTTGTGAGAGCCAGCCCTGTATGCCAACACTGTTGAGCCAGGAGACAGAGCTG 3960  
Qy 3961 TGAGCTCTGCGCTCTTCTGAGCCGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4020  
Db 3961 TGAGCTCTGCGCTCTTCTGAGCCGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4020  
Qy 4021 CCCACTCCCAAGGCTCTACCGAGCTGGGAGGAGTACAGTAGGCTCTGCTCTGCTCTGCTCTG 4080  
Db 4021 CCCACTCCCAAGGCTCTACCGAGCTGGGAGGAGTACAGTAGGCTCTGCTCTGCTCTGCTCTG 4080  
Qy 4081 TTTCTACAGGAAGTCACTCTCGAGGAGTGTGAAGTGTGTGAGTGTGTGAGTGTGTGAGTGTGTG 4140  
Db 4081 TTTCTACAGGAAGTCACTCTCGAGGAGTGTGAAGTGTGTGAGTGTGTGAGTGTGTGAGTGTGTG 4140  
Qy 4141 CATGGCTCTGCTCTTCTGCTTCACTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4200  
Db 4141 CATGGCTCTGCTCTTCTGCTTCACTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4200  
Qy 4201 CATCTGAGGGCAGGGTGTGAGGGCCACACCACTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4260  
Db 4201 CATCTGAGGGCAGGGTGTGAGGGCCACACCACTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4260  
Qy 4261 TTGCTGCCCTCTCTCTCTGAGTGGGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4320  
Db 4261 TTGCTGCCCTCTCTCTCTGAGTGGGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4320  
Qy 4321 TCCAAATTCAGGAAACAAACATGTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4380  
Db 4321 TCCAAATTCAGGAAACAAACATGTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4380  
Qy 4381 CAGGTTGGGCTCTGCAATGAGGCTGCTTAAACAGAGAGAAATATCTGAAAGCAC 4440  
Db 4381 CAGGTTGGGCTCTGCAATGAGGCTGCTTAAACAGAGAGAAATATCTGAAAGCAC 4440





Db 2941 CGTGTGCTTCTCGAACTCCTTTCTTGTCTCTGACATCTCTCTGGGCCACCTTCTA 3000  
Qy 3001 CCTCTCAGCTCTCAGGCTCCTCTTCTCTCTCTCTGCTCTGCCCCACAGGGGCACTCTC 3060  
Db 3001 CCTCTCAGCTCTCAGGCTCCTCTTCTCTCTCTCTGCTCTGCCCCACAGGGGCACTCTC 3060  
Qy 3061 CCAAGTTTGGCCACCCAGCCAAATCAGACAGCTCTCTCTCTGAGGCTCTGCTGCTCTCTC 3120  
Db 3061 CCAAGTTTGGCCACCCAGCCAAATCAGACAGCTCTCTCTCTGAGGCTCTGCTGCTCTCTC 3120  
Qy 3121 CTCTCTCTTTTCTACGCTCTCTCAATGAGAGCTCACACCGCCACTGCTTCAACTGTC 3180  
Db 3121 CTCTCTCTTTTCTAGGCTCTCTCAATGAGAGCTCACACCGCCACTGCTTCAACTGTC 3180  
Qy 3181 ACCTGCATCAAAATGATATCTTATTTGAAAACCTCAGGGAGCCATGACAAAGAGCC 3240  
Db 3181 ACCTGCATCAAAATGATATCTTATTTGAAAACCTCAGGGAGCCATGACAAAGAGCC 3240  
Qy 3241 TAGCATGGAGACAGGCCAGTGTTCAGGGGACACAAAAATAGAAAATTTGGGAGCAGGTA 3300  
Db 3241 TAGCATGGAGACAGGCCAGTGTTCAGGGGACACAAAAATAGAAAATTTGGGAGCAGGTA 3300  
Qy 3301 TCTCTCTGCTGTGAGCCAGCGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3360  
Db 3301 TCTCTCTGCTGTGAGCCAGCGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3360  
Qy 3361 ACAGCTGAATCTGCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGG 3420  
Db 3361 ACAGCTGAATCTGCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGG 3420  
Qy 3421 GGCAGAGCTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGCGAAGCCCTCAATGACAA 3480  
Db 3421 GGCAGAGCTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGCGAAGCCCTCAATGACAA 3480  
Qy 3481 ACTGCGGCTGACCCAGAACTACAGGCTCTACAGCCACTTCTGTGTACTTGGCTGGCT 3540  
Db 3481 ACTGCGGCTGACCCAGAACTACAGGCTCTACAGCCACTTCTGTGTACTTGGCTGGCT 3540  
Qy 3541 CAACTGAGGCTGACCTGCTGAGCTGCGCGAGCTGCGCGAGCTGCGCGAGCTTCTGCAACGCT 3600  
Db 3541 CAACTGAGGCTGACCTGCTGAGCTGCGCGAGCTGCGCGAGCTTCTGCAACGCT 3600  
Qy 3601 CCAGGCTCTGCTGGGAGCAATTCGGGCGTTCATGGGAGCTCTGGGCTTACCCACTGCCCA 3660  
Db 3601 CCAGGCTCTGCTGGGAGCAATTCGGGCGTTCATGGGAGCTCTGGGCTTACCCACTGCCCA 3660  
Qy 3661 GCGCTGCTGGGAGCTGAACCCACTTGGACTGCGCGAGCTGCGCGAGCTGCGCGAGCTTCTCCA 3720  
Db 3661 GCGCTGCTGGGAGCTGAACCCACTTGGACTGCGCGAGCTGCGCGAGCTTCTCCA 3720  
Qy 3721 GAAGATGGAGCACTTGGCTGCTGAGGAGCTGACAGCTGCTGGCTGCGGCTCGGCCAA 3780  
Db 3721 GAAGATGGAGCACTTGGCTGCTGAGGAGCTGACAGCTGCTGGCTGCGGCTCGGCCAA 3780  
Qy 3781 GGAATTTCAACCGGCTCAAGAAGAGTGCAGCTTCCAGAGCTGCAGTCACTGCACTT 3840  
Db 3781 GGAATTTCAACCGGCTCAAGAAGAGTGCAGCTTCCAGAGCTGCAGTCACTGCACTT 3840  
Qy 3841 GGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900  
Db 3841 GGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900  
Qy 3901 CCCACTTTGTAGAGCCAGCCCTGTATGCCAAACCTGTTGAGCCAGGAGCAGAGCTG 3960  
Db 3901 CCCACTTTGTAGAGCCAGCCCTGTATGCCAAACCTGTTGAGCCAGGAGCAGAGCTG 3960  
Qy 3961 TGAGCTCTGGCCCTTTCTGGAACCGCTGGCGGTGTATGCGATCAGCCCTGTCTCTC 4020  
Db 3961 TGAGCTCTGGCCCTTTCTGGAACCGCTGGCGGTGTATGCGATCAGCCCTGTCTCTC 4020  
Qy 4021 CCCACTTCCAAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTGCTCTG 4080  
Db 4021 CCCACTTCCAAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTGCTCTG 4080

Qy 4081 TTTTCTACGAAAGTCTATGCTCGAGGAGTGTGAAGTGTGTAGTGTGTGAGTGTGAGAGCGCT 4140  
Db 4081 TTTTCTACGAAAGTCTATGCTCGAGGAGTGTGAAGTGTGTAGTGTGTGAGTGTGAGAGCGCT 4140  
Qy 4141 CATGGCTCTGCTTCTTGTGCTACCACTTGGCCAGTGCACCCAGCCCTCAGTGTGCA 4200  
Db 4141 CATGGCTCTGCTTCTTGTGCTACCACTTGGCCAGTGCACCCAGCCCTCAGTGTGCA 4200  
Qy 4201 CATCTGAGGAGGAGGCTTGGGGGCCACACACATGCTTCTTCTGGGTGGAAGCCCT 4260  
Db 4201 CATCTGAGGAGGAGGCTTGGGGGCCACACACATGCTTCTTCTGGGTGGAAGCCCT 4260  
Qy 4261 TTGGCTGCCCTCACTCTCTTGTGATGGGTGCTCTTATTCCTTATTCCTCAAAATCACTTATA 4320  
Db 4261 TTGGCTGCCCTCACTCTCTTGTGATGGGTGCTCTTATTCCTTATTCCTCAAAATCACTTATA 4320  
Qy 4321 TCCAAATTCAGGAAACAAACATGCTGGGCAATTTCTACAAAAAGAGATGAGATTAACAGTG 4380  
Db 4321 TCCAAATTCAGGAAACAAACATGCTGGGCAATTTCTACAAAAAGAGATGAGATTAACAGTG 4380  
Qy 4381 CAGGTTGGGCTCTGCAATGGAGTGCCTATAAACCAAGAGAGAAAAATCTGAAAGCAC 4440  
Db 4381 CAGGTTGGGCTCTGCAATGGAGTGCCTATAAACCAAGAGAGAAAAATCTGAAAGCAC 4440  
Qy 4441 AGGGCAGGGACAGACCCAGAGCTCTCCAAAGCACAGAGTGTGAGGTGGAACAAA 4500  
Db 4441 AGGGCAGGGACAGACCCAGAGCTCTCCAAAGCACAGAGTGTGAGGTGGAACAAA 4500  
Qy 4501 ACCGAGCTGAGCATCAGGACCTGCTCCTGAAATTTGTCTTCCAGTATTACGCTCTCTC 4560  
Db 4501 ACCGAGCTGAGCATCAGGACCTGCTCCTGAAATTTGTCTTCCAGTATTACGCTCTCTC 4560  
Qy 4561 TCTGCCCTTTTCCAGGCTATCTGCTGCTTCCAGGCTGCGGAGGCAACCATAGCCAC 4620  
Db 4561 TCTGCCCTTTTCCAGGCTATCTGCTGCTTCCAGGCTGCGGAGGCAACCATAGCCAC 4620  
Qy 4621 ACCACAGGATTTCTGAAAGTTTACAATGCAATGAGTGTGAGGTGGAAGCT 4680  
Db 4621 ACCACAGGATTTCTGAAAGTTTACAATGCAATGAGTGTGAGGTGGAAGCT 4680  
Qy 4681 CCCCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4740  
Db 4681 CCCCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4740  
Qy 4741 TTAATTTTTCAGATGTTATTTATTTAGATGATATTTATTTAGATGATATTTATTTATTTA 4800  
Db 4741 TTAATTTTTCAGATGTTATTTATTTAGATGATATTTATTTAGATGATATTTATTTATTTA 4800  
Qy 4801 TTAACAAATAAAATGCTTGGCCAGAACTTAGTCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCT 4860  
Db 4801 TTAACAAATAAAATGCTTGGCCAGAACTTAGTCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCT 4860  
Qy 4861 TGCTCATCAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920  
Db 4861 TGCTCATCAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920  
Qy 4921 ACAGAGCTCTGGGAGAGGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4980  
Db 4921 ACAGAGCTCTGGGAGAGGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4980  
Qy 4981 CTTACCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5040  
Db 4981 CTTACCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5040  
Qy 5041 TGTCCACCCAGGTTTCTCTCAGCCCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5087  
Db 5041 TGTCCACCCAGGTTTCTCTCAGCCCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5087

RESULT 3

US-09-016-534-3

; Sequence 3, Application US/09016534



Patent No. 6143874  
GENERAL INFORMATION:  
APPLICANT: CHANG, MING-SHI  
APPLICANT: ELLIOTT, GARY S.  
APPLICANT: SARMIENTO, ULLA  
APPLICANT: SENALDI, GIORGIO  
TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMGEN INC.  
STREET: ONE AMGEN CENTER  
CITY: THOUSAND OAKS  
STATE: CA  
COUNTRY: USA  
ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,534  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/792,019  
FILING DATE: 03-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: COOK, ROBERT R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-442B  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5087 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 137..138  
OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED  
OTHER INFORMATION: REGION OF >1KB"  
US-09-016-534-3

Query Match 100.0%; Score 5087; DB 3; Length 5087;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AACCTGCGAGTGGGCTGGCGGATGGATTATTAAAGCTTCGCGGAGCGCGGCTCGCC 60  
Db 1 AACCTGCGAGTGGGCTGGCGGATGGATTATTAAAGCTTCGCGGAGCGCGGCTCGCC 60  
Qy 61 CTCCTCACTCCGCGAGTGGGCTGGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 61 CTCCTCACTCCGCGAGTGGGCTGGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Qy 121 TGGACCTCCGAGCAGGCTTGAACCCAACTAGCCCTGCTTTCATTAACATGACAAGCAG 180  
Db 121 TGGACCTCCGAGCAGGCTTGAACCCAACTAGCCCTGCTTTCATTAACATGACAAGCAG 180  
Qy 181 CGCCCCATCTGATACCTAAACCGAGCAAGTCAACGCTCAACGCTCAACGCTCAACGCT 240  
Db 181 CGCCCCATCTGATACCTAAACCGAGCAAGTCAACGCTCAACGCTCAACGCTCAACGCT 240  
Qy 241 CAGACCTCACCACATCTTGTGGATCAAACTCAACGCTCAACGCTCAACGCTCAACGCT 300  
Db 241 CAGACCTCACCACATCTTGTGGATCAAACTCAACGCTCAACGCTCAACGCTCAACGCT 300  
Qy 301 AGTCTAACTAATCTGAACCTTTTAAAGTAAACCCAGTCCCTTAAACCTAAGCCCAAT 360  
Db 301 AGTCTAACTAATCTGAACCTTTTAAAGTAAACCCAGTCCCTTAAACCTAAGCCCAAT 360

Qy 361 GCCAATTATATCTACCTAGCCAAACCTTAACCTGCTTTCGCGAGTCCAAAGTGTCCACTG 420  
Db 361 GCCAATTATATCTACCTAGCCAAACCTTAACCTGCTTTCGCGAGTCCAAAGTGTCCACTG 420  
Qy 421 AATCCTCACCTTGGTCTCTCACTGAAATCCAGAAAAGCATATTTCCGCCACTGCGCCACAT 480  
Db 421 AATCCTCACCTTGGTCTCTCACTGAAATCCAGAAAAGCATATTTCCGCCACTGCGCCACAT 480  
Qy 481 CCTCCTTACAGCACCCCAACCTTGGCTCTGGACTCTCTGGTATCTCTGGGATGTCCAAACT 540  
Db 481 CCTCCTTACAGCACCCCAACCTTGGCTCTGGACTCTCTGGTATCTCTGGGATGTCCAAACT 540  
Qy 541 CTGAGTGCCTCAGCCCAACCAAGCCGACTCGTCAATGACACCTCTCTCCCTTCTCTCTCC 600  
Db 541 CTGAGTGCCTCAGCCCAACCAAGCCGACTCGTCAATGACACCTCTCTCCCTTCTCTCTCC 600  
Qy 601 CACCTTTCAGGCTGATGGAAGGCTCATTTGAAGTCCAACTTTTCCCCACCTTAAACACC 660  
Db 601 CACCTTTCAGGCTGATGGAAGGCTCATTTGAAGTCCAACTTTTCCCCACCTTAAACACC 660  
Qy 661 AAGAACGGGGTGAACCTCCACACTGCGACCGTTCCCTGAGAGTGAGCACTAAATCTCCTT 720  
Db 661 AAGAACGGGGTGAACCTCCACACTGCGACCGTTCCCTGAGAGTGAGCACTAAATCTCCTT 720  
Qy 721 CAATCTAACCCCACTTACACTTCCCACTCAGGAATCACATCTAGAAATATACCCAAA 780  
Db 721 CAATCTAACCCCACTTACACTTCCCACTCAGGAATCACATCTAGAAATATACCCAAA 780  
Qy 781 ACTAGCCCAATAAGGACGCGGACCTAGTGTCTTAACCTATACCTTCTCTTCTCTATGG 840  
Db 781 ACTAGCCCAATAAGGACGCGGACCTAGTGTCTTAACCTATACCTTCTCTTCTCTATGG 840  
Qy 841 GTGAGTCTGTCTTGGCGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
Db 841 GTGAGTCTGTCTTGGCGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
Qy 901 AGCTGCCAGCTCTGACATGTGTCTCTCCCACTCTGACTCTGACTCTGACTCTGACTCTG 960  
Db 901 AGCTGCCAGCTCTGACATGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960  
Qy 961 GACTGGAGACTGGCAGGAGCTAGGATCAATGTGAAACACAGGAGGCTGAGCTGAGTGCAGT 1020  
Db 961 GACTGGAGACTGGCAGGAGCTAGGATCAATGTGAAACACAGGAGGCTGAGCTGAGTGCAGT 1020  
Qy 1021 CCTTAGGCTGGCCCGCTCCCTCCATGTACACATATATATATATATATATATATATATAT 1080  
Db 1021 CCTTAGGCTGGCCCGCTCCCTCCATGTACACATATATATATATATATATATATATAT 1080  
Qy 1081 GCACATGTGCCAAAGACTCTCTCAGCTGACACACAGATCCATTTCTCAAGTATCTACTGAT 1140  
Db 1081 GCACATGTGCCAAAGACTCTCTCAGCTGACACACAGATCCATTTCTCAAGTATCTACTGAT 1140  
Qy 1141 AGACACTCATGGTGGCAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200  
Db 1141 AGACACTCATGGTGGCAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200  
Qy 1201 GTCTTGGCAGGAGTGTCT 1260  
Db 1201 GTCTTGGCAGGAGTGTCT 1260  
Qy 1261 TCACCCCAACCCAGGCTGGGAGCAGACACTTGGGGGCTGCGGAGTCTCTCTCTCTCTCT 1320  
Db 1261 TCACCCCAACCCAGGCTGGGAGCAGACACTTGGGGGCTGCGGAGTCTCTCTCTCTCTCT 1320  
Qy 1321 GTGTGGGCGCGGCGGCTCATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380  
Db 1321 GTGTGGGCGCGGCGGCTCATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380  
Qy 1381 GATGTAGCTGCTGTGCAAGGCTCTGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440  
Db 1381 GATGTAGCTGCTGTGCAAGGCTCTGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440  
Qy 1441 CACAGGGAGCCAGGGGCTGGCCCTCCATCCAGAAAACCTTATGACCTTACCCCGCTACCT 1500



QY 3661 GCCGCTGCTGGGACTGAACCCACTTTGGACTCTCTGGCCCTGGCCACAGTGAATCTCTCCA 3720  
Db 3661 GCCGCTGCTGGGACTGAACCCACTTTGGACTCTCTGGCCCTGGCCACAGTGAATCTCTCCA 3720  
QY 3721 GAAGATGAGACGACTTCTGGCTGCTGAAGAGCTGACAGCTGGCTGTGGCTCGGCCAA 3780  
Db 3721 GAAGATGAGACGACTTCTGGCTGCTGAAGAGCTGACAGCTGGCTGTGGCTCGGCCAA 3780  
QY 3781 GGACTTCAACCGGCTCAAGAAGAGATGACAGCTCCAGCAGTGCAGTCAACCTGCACCT 3840  
Db 3781 GGACTTCAACCGGCTCAAGAAGAGATGACAGCTCCAGCAGTGCAGTCAACCTGCACCT 3840  
QY 3841 GGGGGCTATGGCTTCTGACTTCTGACCTTCTCTCTTCCGCTCCGCTTCAAAACCTGCT 3900  
Db 3841 GGGGGCTATGGCTTCTGACTTCTGACCTTCTCTCTTCCGCTCCGCTTCAAAACCTGCT 3900  
QY 3901 CCCACTTTGTGAGAGCCAGCCCTGTATGCCAACACCTGTGTGAGCCAGGAGCAGAGCTG 3960  
Db 3901 CCCACTTTGTGAGAGCCAGCCCTGTATGCCAACACCTGTGTGAGCCAGGAGCAGAGCTG 3960  
QY 3961 TGAGCTCTGGCCCTTTCTGACCGGCTGGGCTGGCTGATGCGATCAGCCCTGTCTCCTC 4020  
Db 3961 TGAGCTCTGGCCCTTTCTGACCGGCTGGGCTGGCTGATGCGATCAGCCCTGTCTCCTC 4020  
QY 4021 CCCACTTCCAAAGTCTTACCGAGCTGGGGAGGAGTACAGTAGGCCCTGTCTCTCTG 4080  
Db 4021 CCCACTTCCAAAGTCTTACCGAGCTGGGGAGGAGTACAGTAGGCCCTGTCTCTCTG 4080  
QY 4081 TTTCTACAGGAGTCACTCTCCAGGAGTGAAGTGTGAGTGTGAGTGTGAGGAGGCT 4140  
Db 4081 TTTCTACAGGAGTCACTCTCCAGGAGTGAAGTGTGAGTGTGAGTGTGAGGAGGCT 4140  
QY 4141 CATGCCCTCTGCTTTCTTGCTTACCACCTTGGCCAGTGGCCACCCAGCCCTCAGTGGCA 4200  
Db 4141 CATGCCCTCTGCTTTCTTGCTTACCACCTTGGCCAGTGGCCACCCAGCCCTCAGTGGCA 4200  
QY 4201 CATCTGGAGGGCAGGGTGTAGGGGCCACCAACACATGCTTTCTGGGGTGAAGCCCT 4260  
Db 4201 CATCTGGAGGGCAGGGTGTAGGGGCCACCAACACATGCTTTCTGGGGTGAAGCCCT 4260  
QY 4261 TTGGCTGCCCACTCTCTTGTGATGGGTGTCTCCCTTATCCCAATCACTCTATACA 4320  
Db 4261 TTGGCTGCCCACTCTCTTGTGATGGGTGTCTCCCTTATCCCAATCACTCTATACA 4320  
QY 4321 TCCAAATCAGGAAACAAATGTTGGCAATCTACACAAAAGAGATGAGATTACAGTG 4380  
Db 4321 TCCAAATCAGGAAACAAATGTTGGCAATCTACACAAAAGAGATGAGATTACAGTG 4380  
QY 4381 CAGGGTTGGGTCTGCATTTGAGGTGCCCTATAAACAGAGAGAAATATCTGAAGCAC 4440  
Db 4381 CAGGGTTGGGTCTGCATTTGAGGTGCCCTATAAACAGAGAGAAATATCTGAAGCAC 4440  
QY 4441 AGGGCAGGGACAGACACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAA 4500  
Db 4441 AGGGCAGGGACAGACACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAA 4500  
QY 4501 ACCGAGCTGACATCAGGACCTTGGCTCGAATTTCTTCCAGTATTACGGTGCCTCTTC 4560  
Db 4501 ACCGAGCTGACATCAGGACCTTGGCTCGAATTTCTTCCAGTATTACGGTGCCTCTTC 4560  
QY 4561 TCTGCCCCCTTCCAGGGTATCTGTGGGTTCAGGGTCCAGGCTGGGGAGGCAACATAGCCAC 4620  
Db 4561 TCTGCCCCCTTCCAGGGTATCTGTGGGTTCAGGGTTCAGGCTGGGGAGGCAACATAGCCAC 4620  
QY 4621 ACCACAGGATTTCTGAAAGTTTACAATGACAGTACATTTTGGGGTGTAGGGTGCAGCT 4680  
Db 4621 ACCACAGGATTTCTGAAAGTTTACAATGACAGTACATTTTGGGGTGTAGGGTGCAGCT 4680  
QY 4681 CCCCAAGCCCTGCCCCCAGCCCCCAGCCCACTCATGACTCTAAGTGTCTGTATTATAT 4740  
Db 4681 CCCCAAGCCCTGCCCCCAGCCCCCAGCCCACTCATGACTCTAAGTGTCTGTATTATAT 4740

QY 4741 TTATTATTGGAGATGTTATTATTATAGATGATATTATTGAGAAATTTCTATTCTGTA 4800  
Db 4741 TTATTATTGGAGATGTTATTATTATAGATGATATTATTGAGAAATTTCTATTCTGTA 4800  
QY 4801 TTAAACAATAAATGTTGCCCCAGAACTTAGTCTCTTTTGGCCAGCTCACCCCTCCTGG 4860  
Db 4801 TTAAACAATAAATGTTGCCCCAGAACTTAGTCTCTCTTTGGCCAGCTCACCCCTCCTGG 4860  
QY 4861 TGCTCATCAGACTTGTGCCACCCCTGGCTCCACCTCCCTGCTTGTCTGTGGAGCTGC 4920  
Db 4861 TGCTCATCAGACTTGTGCCACCCCTGGCTCCACCTCCCTGCTTGTCTGTGGAGCTGC 4920  
QY 4921 ACAGAGCTCTGGAAAGAGCCCTCTTCTCCCGCACTGGGGCGATGGGCGCACCTCACA 4980  
Db 4921 ACAGAGCTCTGGAAAGAGCCCTCTTCTCCCGCACTGGGGCGATGGGCGCACCTCACA 4980  
QY 4981 CTTACCCACTGCTGTGCCACCAACCAACCCCTTGTATCCCTCAGTCTCCACAGCTTC 5040  
Db 4981 CTTACCCACTGCTGTGCCACCAACCAACCCCTTGTATCCCTCAGTCTCCCTCAGCTTC 5040  
QY 5041 TGTCACCCAGGTTTCCCTCACCACCCCTTTGCTAAAGTCTTCCCTCA 5087  
Db 5041 TGTCACCCAGGTTTCCCTCACCACCCCTTTGCTAAAGTCTTCCCTCA 5087

## RESULT 4

US-09-106-182-1  
; Sequence 1, Application US/09106182  
; Patent No. 6046035  
; GENERAL INFORMATION:  
; APPLICANT: Shi, Yanggu  
; APPLICANT: Ruben, Steve  
; TITLE OF INVENTION: Cardiostrophin-Like Cytokine  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc  
; STREET: 9410 Key West Ave  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: US  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/09/106,182  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/051,053  
; FILING DATE: 30-JUN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PP385  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1710 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 46..720  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 46..126

FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 127..720  
US-09-106-182-1

Query Match 28.8%; Score 1463.4; DB 3; Length 1710;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1464; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	3363	ACGTGAACCTACCTGGCCCGCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGGCTGGGG	3422
DB	227	ATCTGAACCTACCTGGCCCGCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGGCTGGGG	286
QY	3423	CAGAGACTCTGCCCGGCGCACTGTTGACTTGGAGGTGGCGAAGGCTCAATGACAAAC	3482
DB	287	CAGAGACTCTGCCCGGCGCACTGTTGACTTGGAGGTGGCGAAGGCTCAATGACAAAC	346
QY	3483	TGCGGTGACCCAGAACTACGAGGCTTACAGCCACCTTCTGTGTTACTTCGCTGGCTCA	3542
DB	347	TGCGGTGACCCAGAACTACGAGGCTTACAGCCACCTTCTGTGTTACTTCGCTGGCTCA	406
QY	3543	ACGTCAGGCTGCCACTGCTAGCTGGCGGCGAGCCTGGCCCACTTCTGACACAGCCTCC	3602
DB	407	ACGTCAGGCTGCCACTGCTAGCTGGCGGCGAGCCTGGCCCACTTCTGACACAGCCTCC	466
QY	3603	AGGCGCTGCTGGCGAGCACTTGGCGGCGTCTATGCGAGCTCTGGGCTACCCACTGCCCCAGC	3662
DB	467	AGGCGCTGCTGGCGAGCACTTGGCGGCGTCTATGCGAGCTCTGGGCTACCCACTGCCCCAGC	526
QY	3663	CGTGTGCTGGGACTGAACCCACTTGGACTCTCGCCCTGCCACAGTGACTTCTCCAGA	3722
DB	527	CGTGTGCTGGGACTGAACCCACTTGGACTCTCGCCCTGCCACAGTGACTTCTCCAGA	586
QY	3723	AGATGACGACTTCTGGCTGCTAGGAGCTGCAGACCTGGCTGGGCTCGGCCAAGG	3782
DB	587	AGATGACGACTTCTGGCTGCTAGGAGCTGCAGACCTGGCTGGGCTCGGCCAAGG	646
QY	3783	ACTTCAACCGGCTCAAGAAAGATGAGGCTCCAGCAGCTGCACTGCACTGCACTGCG	3842
DB	647	ACTTCAACCGGCTCAAGAAAGATGAGGCTCCAGCAGCTGCACTGCACTGCACTGCG	706
QY	3843	GGGCTCATGGCTTCTGACTTCTGACTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT	3902
DB	707	GGGCTCATGGCTTCTGACTTCTGACTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT	766
QY	3903	CACCTTGTGAGAGCAGCCCTGTATGCAACCTGTTGAGCCAGGAGAGAGAGCTGTG	3962
DB	767	CACCTTGTGAGAGCAGCCCTGTATGCAACCTGTTGAGCCAGGAGAGAGAGCTGTG	826
QY	3963	AGCCTCTGGCCCTTCTGACCGGCTGGGCTGTGATGCGATCAGCCCTGCTCCTCCC	4022
DB	827	AGCCTCTGGCCCTTCTGACCGGCTGGGCTGTGATGCGATCAGCCCTGCTCCTCCC	886
QY	4023	CACCTCCCAAGGCTTACCGAGCTGGGAGGAGTACAGTAGGCGCTGCTGCTGCTGT	4082
DB	887	CACCTCCCAAGGCTTACCGAGCTGGGAGGAGTACAGTAGGCGCTGCTGCTGCTGT	946
QY	4083	TCTACAGGAAGTCTGCTCAGGAGTGTGAGTGGTTCAGTGGTTCAGAGGCGCTCA	4142
DB	947	TCTACAGGAAGTCTGCTCAGGAGTGTGAGTGGTTCAGTGGTTCAGAGGCGCTCA	1006
QY	4143	TGGCTCTGCTTCTTCCCTACCACTTGGCGAGTGGCCAGCCAGCCCTCAGGTGGCACA	4202
DB	1007	TGGCTCTGCTTCTTCCCTACCACTTGGCGAGTGGCCAGCCAGCCCTCAGGTGGCACA	1066
QY	4203	TCTGGAGGCGAGGGGTTGAGGGGCCACACACATGCGCTTCTGGGGTGAAGCCCTTT	4262
DB	1067	TCTGGAGGCGAGGGGTTGAGGGGCCACACACATGCGCTTCTGGGGTGAAGCCCTTT	1126
QY	4263	GGCTGCCCACTCTCCCTGGATGGGTGTGCTCCCTTATCCCAATCAGCTCTATACATC	4322
DB	1127	GGCTGCCCACTCTCCCTGGATGGGTGTGCTCCCTTATCCCAATCAGCTCTATACATC	1186

QY	4323	CAATTCCAGGAAACAAACATGGTGGCAATTTCTACACAAAGAGAGATGAGATTAACTGCA	4382
DB	1187	CAATTCCAGGAAACAAACATGGTGGCAATTTCTACACAAAGAGAGATGAGATTAACTGCA	1246
QY	4383	GGGTTGGGCTCTGCAATTTGGAGGTGCCCTATATAACAGAGAGAAATACTGAAAGCACAG	4442
DB	1247	GGGTTGGGCTCTGCAATTTGGAGGTGCCCTATATAACAGAGAGAAATACTGAAAGCACAG	1306
QY	4443	GGGTCAGGAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG	4502
DB	1307	GGGTCAGGAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG	1366
QY	4503	CCGAGCTGAGCATCAGGACCTTGGCTCGAATTTGCTTCCAGTATTAACGTCCTTCTTC	4562
DB	1367	CCGAGCTGAGCATCAGGACCTTGGCTCGAATTTGCTTCCAGTATTAACGTCCTTCTTC	1426
QY	4563	TGCCCCCTTTCCAGGGTATCTGTGGGTGCGAGCTCGGGAGGCAACCATAGCCACAC	4622
DB	1427	TGCCCCCTTTCCAGGGTATCTGTGGGTGCGAGCTCGGGAGGCAACCATAGCCACAC	1486
QY	4623	CACAGGATTTCTGAAAGTTTACCAATGCAATGCAATGCAATTTGGGGTGGAGCTCC	4682
DB	1487	CACAGGATTTCTGAAAGTTTACCAATGCAATGCAATGCAATTTGGGGTGGAGCTCC	1546
QY	4683	CCAAAGGCTGCCCCCAGCCCCCAGCCCCCAGCCCCCAGCCCCCAGCCCCCAGCCCCC	4742
DB	1547	CCAAAGGCTGCCCCCAGCCCCCAGCCCCCAGCCCCCAGCCCCCAGCCCCCAGCCCCC	1606
QY	4743	ATTTATTTGGAGATGTTATTTATTTAGATGATATTTATTTAGCAGATTTCTTCTGATT	4802
DB	1607	ATTTATTTGGAGATGTTATTTATTTAGATGATATTTATTTAGCAGATTTCTTCTGATT	1666
QY	4803	AACAAATAAAATGCTTGCCCCAGAA	4827
DB	1667	AACAAATAAAATGCTTGCCCCAGAA	1691

## RESULT 5

US-08-792-019B-1

; Sequence 1, Application US/08792019B

; Patent No. 5741772

; GENERAL INFORMATION:

; APPLICANT: CHANG, MING-SHI

; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: AMGEN INC.

; STREET: 1840 DEHAVILLAND DRIVE

; CITY: THOUSAND OAKS

; STATE: CA

; COUNTRY: USA

; ZIP: 91320

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/792,019B

; FILING DATE: 03-FEB-1997

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: COOK, ROBERT R.

; REGISTRATION NUMBER: 31,602

; REFERENCE/DOCKET NUMBER: A-442

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 797 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

```
; NAME/KEY: CDS
; LOCATION: 90...764
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 171...764
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90...170
US-08-792-019B-1

Query Match
Best Local Similarity 10.3%; Score 523.4; DB 1; Length 797;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422
Db 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 330

QY 3423 CAGAGACTCTGCCAGGCCACTTGTGACTTGGAGGTGGGAGCCCTCAATGACAAAC 3482
Db 331 CAGAGACTCTGCCAGGCCACTTGTGACTTGGAGGTGGGAGCCCTCAATGACAAAC 390

QY 3483 TCGGCTGACCCAGAACTACGAGGCTTACAGCCACCTTCTGTGTTACTTGGGTGGCTCA 3542
Db 391 TCGGCTGACCCAGAACTACGAGGCTTACAGCCACCTTCTGTGTTACTTGGGTGGCTCA 450

QY 3543 ACCGTGAGCTGCGACTGCTGAGCTGCGCGCAGCCCTGCGCCACTTCTGCACGAGCTCC 3602
Db 451 ACCGTGAGCTGCGACTGCTGAGCTGCGCGCAGCCCTGCGCCACTTCTGCACGAGCTCC 510

QY 3603 AGGGCTGCTGGGCGAGCTTGGGGGCTCATGGCAGCTTGGGCTACCCACTGCCCCCAGC 3662
Db 511 AGGGCTGCTGGGCGAGCTTGGGGGCTCATGGCAGCTTGGGCTACCCACTGCCCCCAGC 570

QY 3663 CGCTGCTGGGACTGAACCCACTTGGACTTGGCCCTGCGCCAGCTTCTCTCCCGA 3722
Db 571 CGCTGCTGGGACTGAACCCACTTGGACTTGGCCCTGCGCCAGCTTCTCTCCCGA 630

QY 3723 AGATGGAGCTTCTGCTGCTGAAGAGCTGACAGCTGCTGGGCTGCGGCTCGGCAAGG 3782
Db 631 AGATGGAGCTTCTGCTGCTGAAGAGCTGACAGCTGCTGGGCTGCGGCTCGGCAAGG 690

QY 3783 ACTTCAACCGGCTCAAGAAGAGTGCAGCTTCCAGAGCTGCGAGCTGCGAGCTGCGAGCTGG 3842
Db 691 ACTTCAACCGGCTCAAGAAGAGTGCAGCTTCCAGAGCTGCGAGCTGCGAGCTGCGAGCTGG 750

QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCTGCTCCCC 3887
Db 751 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCTGCTCCCC 795
```

## RESULT 6

```
US-08-988-819-1
; Sequence 1, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUTROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
```

```
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90...764
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 171...764
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90...170
US-08-988-819-1
```

## Query Match

```
Best Local Similarity 10.3%; Score 523.4; DB 3; Length 797;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422
Db 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 330

QY 3423 CAGAGACTCTGCCAGGCCACTTGTGACTTGGAGGTGGGAGCCCTCAATGACAAAC 3482
Db 331 CAGAGACTCTGCCAGGCCACTTGTGACTTGGAGGTGGGAGCCCTCAATGACAAAC 390

QY 3483 TCGGCTGACCCAGAACTACGAGGCTTACAGCCACCTTCTGTGTTACTTGGGTGGCTCA 3542
Db 391 TCGGCTGACCCAGAACTACGAGGCTTACAGCCACCTTCTGTGTTACTTGGGTGGCTCA 450

QY 3543 ACCGTGAGCTGCGACTGCTGAGCTGCGCGCAGCCCTGCGCCACTTCTGCACGAGCTCC 3602
Db 451 ACCGTGAGCTGCGACTGCTGAGCTGCGCGCAGCCCTGCGCCACTTCTGCACGAGCTCC 510

QY 3603 AGGGCTGCTGGGCGAGCTTGGGGGCTCATGGCAGCTTGGGCTACCCACTGCCCCCAGC 3662
Db 511 AGGGCTGCTGGGCGAGCTTGGGGGCTCATGGCAGCTTGGGCTACCCACTGCCCCCAGC 570

QY 3663 CGCTGCTGGGACTGAACCCACTTGGACTTGGCCCTGCGCCAGCTTCTCTCCCGA 3722
Db 571 CGCTGCTGGGACTGAACCCACTTGGACTTGGCCCTGCGCCAGCTTCTCTCCCGA 630

QY 3723 AGATGGAGCTTCTGCTGCTGAAGAGCTGACAGCTGCTGGGCTGCGGCTCGGCAAGG 3782
Db 631 AGATGGAGCTTCTGCTGCTGAAGAGCTGACAGCTGCTGGGCTGCGGCTCGGCAAGG 690

QY 3783 ACTTCAACCGGCTCAAGAAGAGTGCAGCTTCCAGAGCTGCGAGCTGCGAGCTGCGAGCTGG 3842
Db 691 ACTTCAACCGGCTCAAGAAGAGTGCAGCTTCCAGAGCTGCGAGCTGCGAGCTGCGAGCTGG 750

QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCTGCTCCCC 3887
Db 751 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCTGCTCCCC 795
```

## RESULT 7

```
US-09-016-534-1
; Sequence 1, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
```

APPLICANT: CHANG, MING-SHI  
APPLICANT: ELLIOTT, GARY S.  
APPLICANT: SARMIENTO, ULLA  
APPLICANT: SENALDI, GIORGIO  
TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMGEN INC.  
STREET: ONE AMGEN CENTER  
CITY: THOUSAND OAKS  
STATE: CA  
COUNTRY: USA  
ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,534  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/792,019  
FILING DATE: 03-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: COOK, ROBERT R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-442B  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 797 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 90..764  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 171..764  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 90..170  
US-09-016-534-1

Query Match 10.3%; Score 523.4; DB 3; Length 797;  
Best Local Similarity 99.8%; Pred. No. 1.1e-120;  
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGG 3422  
DB 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGG 330  
QY 3423 CAGAGACTCTCCCGAGGCCACTGTGTGACTTGGAGGTGGCGAGCCCTCAATGACAAAC 3482  
DB 331 CAGAGACTCTCCCGAGGCCACTGTGTGACTTGGAGGTGGCGAGCCCTCAATGACAAAC 390  
QY 3483 TGGCGGTGACCCAGAACTACAGAGGCCCTACGCCACTTCTGTGTTACTTGGCGTGCCTCA 3542  
DB 391 TGGCGGTGACCCAGAACTACAGAGGCCCTACGCCACTTCTGTGTTACTTGGCGTGCCTCA 450  
QY 3543 ACCGTGAGGCTGACCTCTGAGCTGCGCGAGCCCTGCGCCACTTCTGACACAGCCCTCC 3602  
DB 451 ACCGTGAGGCTGACCTCTGAGCTGCGCGAGCCCTGCGCCACTTCTGACACAGCCCTCC 510  
QY 3603 AGGCGCTCTGGGAGCAATTCGGCGCTGATGGCGAGCTCTGGGCTACCCACTTGGCCCGAGC 3662  
DB 511 AGGCGCTCTGGGAGCAATTCGGCGCTGATGGCGAGCTCTGGGCTACCCACTTGGCCCGAGC 570  
QY 3663 CGCTGCCTGGGACTGAAACCCACTTGGACTCTCTGGCCCTGCCACAGTGAATCTCTCCACGA 3722

## RESULT 8

US-08-792-019B-4  
Sequence 4, Application US/08792019B  
Patent No. 5741772  
GENERAL INFORMATION:

APPLICANT: CHANG, MING-SHI  
TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMGEN INC.  
STREET: 1840 DEHAVILLAND DRIVE  
CITY: THOUSAND OAKS  
STATE: CA  
COUNTRY: USA  
ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/792,019B  
FILING DATE: 03-FEB-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: COOK, ROBERT R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-442

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 819 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 95..769  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 176..769  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 95..175  
US-08-792-019B-4

Query Match 8.4%; Score 429.6; DB 1; Length 819;  
Best Local Similarity 90.4%; Pred. No. 2.3e-97;  
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGG 3422  
DB 276 ACCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGG 335  
QY 3423 CAGAGACTCTGCCAGGCCACTTGTGACTTGGAGGTGGCGAGCCCTCAATGACAAAC 3482  
DB 336 CAGAAACTCTGCCAGGCCACTTGTGAAAGTGTGGCGAAGCCCTCAATGACAGGC 395

QY 3483 TCGGCTGACCCAGCAACTAGAGGCGCTACAGCCACCTTCTGTGTACTTGTGCGGCTCA 3542  
Db 396 TCGGCTGACCCAGCAACTATGAGGCGTACAGTCACTCTCTGTGTACTTGTGCGGCTCA 455  
QY 3543 ACCGTGAGCTGCCAGCTGCTGAGCTGCGCGGAGCTGCGCCACTTCTGTGACAGGCTCC 3602  
Db 456 ACCGTGAGCTGCCAGCTGCTGAGCTGCGCGGAGCTGCGCCACTTCTGTGACAGGCTCC 515  
QY 3603 AGGGCTGCTGGGCGCACTTGGGGCGTATGGCGAGCTTGGGCTACCCACTGCCCCCAGC 3662  
Db 516 AGGGCTGCTGGGCGCACTTGGCGAGCTTGGCGAGCTTGGGCTACCCACTGCCCCCAGC 575  
QY 3663 CGCTGCTGGGAGTGAACCCACTTGGATCTTGGGCGCTGCGCCACTGCTTCTCTCCAGA 3722  
Db 576 CTCTCCAGGAGTGAACCCACTTGGCGAGCTTGGGCGCTGCGCCACTGCTTCTCTCCAGA 635  
QY 3723 AGATGAGCACTTCTGGCTGCTGAAGGAGCTGCAGAGCTGCGGCTGCGGCGGCAAGG 3782  
Db 636 AGATGAGCACTTCTGGCTGCTGAAGGAGCTGCAGAGCTGCGGCTGCGGCGGCAAGG 695  
QY 3783 ACTTCAACCGGCTCAAGAGAGATGCGAGCTTCCAGAGCTGCGAGTCACTTCCAGCTGG 3842  
Db 696 ACTTCAACCGGCTTAAAGAGAGATGCGAGCTTCCAGAGCTTCCAGAGCTTCCAGCTGG 755  
QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870  
Db 756 AGGCACATGGTTTCTGACCTCTGACCTT 783

## RESULT 9

US-09-988-819-4  
; Sequence 4, Application US/08988819  
; Patent No. 6054294  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/988,819  
; FILING DATE: 12-DEC-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442A  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 819 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 95..769  
; FEATURE:  
; NAME/KEY: mat peptide  
; LOCATION: 176..769

## ; FEATURE:

; NAME/KEY: sig\_peptide  
; LOCATION: 95..175  
US-08-988-819-4

Query Match 8.4%; Score 429.6; DB 3; Length 819;  
Best Local Similarity 90.4%; Pred. No. 2.3e-97;  
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACTCTGGGCCCCCTTTCAAGAGCCAGACTTCAACCTCTCCCGCTGGGGG 3422  
Db 276 ACCTGAACCTACTCTGGGCCCCCTTTCAAGAGCCAGACTTCAATCTCTCGACTGGGGG 335  
QY 3423 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGGAGAGCTCAATGACAAAC 3482  
Db 336 CAGAAACTCTGCCAGGGCCAGGTCAACTTGGAGGTGTGGGAGAGCTCAATGACAGGC 395  
QY 3483 TCGGCTGACCCAGCAACTACAGAGGCTTACAGCCACTTCTGTGTACTTGTGCGGCTCA 3542  
Db 396 TCGGCTGACCCAGCAACTATGAGGCGTACAGTCACTCTCTGTGTACTTGTGCGGCTCA 455  
QY 3543 ACCGTGAGCTGCCAGCTGCTGAGCTGCGCGGAGCTGCGCCACTTCTGACAGGCTCC 3602  
Db 456 ACCGTGAGCTGCCAGCTGCTGAGCTGCGCGGAGCTGCGCCACTTCTGACAGGCTCC 515  
QY 3603 AGGGCTGCTGGGCGAGCACTTGGGCGCTCATGGCAGCTTGGGCTACCCACTGCCCCCAGC 3662  
Db 516 AGGGCTGCTGGGCGAGCACTTGGCAGGTGTATGGCGAGCTTGGCTACCCACTGCCCCCAGC 575  
QY 3663 CGTGCCTGGGAGTGAACCCACTTGGACTCTCTGGGCGCTGCGCCACTGAGTGACTTCTCCAGA 3722  
Db 576 CTCTGCCAGGAGTGAACCCACTTGGGCGCTGCGGCGCTGCGCCACTGAGTGACTTCTCCAGA 635  
QY 3723 AGATGAGCACTTCTGGCTGCTGAAGAGCTGCGAGAGCTGCGAGCTGCGGCTGCGGCTCAAGG 3782  
Db 636 AGATGAGCACTTCTGGCTGCTGAAGAGAGCTGCGAGAGCTGCGAGCTGCGGCTTTCAGGCAAGG 695  
QY 3783 ACTTCAACCGGCTCAAGAGAGATGCGAGCTTCCAGAGCTGCGAGTCACTTCCAGCTGG 3842  
Db 696 ACTTCAACCGGCTTAAAGAGAGATGCGAGCTTCCAGAGCTTCCAGAGCTTCCAGCTGG 755  
QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870  
Db 756 AGGCACATGGTTTCTGACCTCTGACCTT 783

## RESULT 10

US-09-016-534-4  
; Sequence 4, Application US/09016534  
; Patent No. 6143874  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; APPLICANT: ELLIOTT, GARY S.  
; APPLICANT: SARMIENTO, ULLA  
; APPLICANT: SENALDI, GIORGIO  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,534  
; FILING DATE:  
; CLASSIFICATION:

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442B  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 819 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 95..769  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 176..769  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 95..175  
; US-09-016-534-4

Query Match 8.4%; Score 429.6; DB 3; Length 819;  
Best Local Similarity 90.4%; Pred. No. 2.3e-97; Indels 0; Gaps 0;  
Matches 459; Conservative 0; Mismatches 49;  
QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 3422  
DB 276 ACCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCTGACTGGGG 335  
QY 3423 CAGAGACTCTGCCAGGCGCACTGTGACTTGGAGGTGTGGAGGCTCAATGACAAAC 3482  
DB 336 CAGAAACTCTGCCAGGCGCACTGTGACTTGGAGGTGTGGAGGCTCAATGACAGGC 395  
QY 3483 TGGCGGTGACCCAGAACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGGCTCA 3542  
DB 396 TGGCGGTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTGTTACTTGGTGGCTCA 455  
QY 3543 ACCGTGAGGCTGCACTGTGAGTGGCGCGCGAGCCCTGGCCACTTTGCAACGACCTCC 3602  
DB 456 ACCGTGAGGCTGCACTGTGAGTGGCGCGCGAGCCCTGGCCACTTTGCAACGACCTCC 515  
QY 3603 AGGCGCTGTGGGAGCAATTCGGGCGCTCATGGAGCTCTGGGCTACCACTGCCCGAGC 3662  
DB 516 AGGCGCTGTGGGAGCAATTCGGGCGCTCATGGAGCTCTGGGCTACCACTGCCCGAGC 575  
QY 3663 CGTGTGCTGGGACTGAACCCCACTTGGGACTCTGCGCCCTGCGCCAGAGTGAATTCCTCCAGA 3722  
DB 576 CTCTGCCAGGACTGAGCCAGCCCTGGGCGCTGCGCCCTGCGCCAGAGTGAATTCCTCCAGA 635  
QY 3723 AGATGAGCACTTCTGGCTGCTGAAGAGTGTGAGAGCTGAGAGCTGCGCTGCGGCTCGGCAAGG 3782  
DB 636 AGATGAGCACTTCTGGCTGCTGAAGAGTGTGAGAGCTGAGAGCTGCGCTGCGGCTCGGCAAGG 695  
QY 3783 ACTTCAACCGGCTCAAGAGAGAGTGAAGCTTCCAGAGCTGCACTGCACTGCACTGCACTGG 3842  
DB 696 ACTTCAACCGGCTTAAAGAGAGATGAGAGCTTCCAGAGCTGCACTGCACTGCACTGCACTGG 755  
QY 3843 GGGCTCATGCTTCTGACTTCTGACCTT 3870  
DB 756 AGGCATATGGTCTTCTGACCTCTGACCT 783

RESULT 11  
US-09-106-182-7  
; Sequence 7, Application US/09106182  
; Patent No. 6046035  
; GENERAL INFORMATION:  
; APPLICANT: Shi, Yanguu  
; APPLICANT: Ruben, Steve

TITLE OF INVENTION: Cardiostrophin-Like Cytokine  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc  
; STREET: 9410 Key West Ave  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: US  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/106,182  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/051,053  
; FILING DATE: 30-JUN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PP385  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 396 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-09-106-182-7

Query Match 6.4%; Score 324.6; DB 3; Length 396;  
Best Local Similarity 95.2%; Pred. No. 2.1e-71;  
Matches 374; Conservative 0; Mismatches 15; Indels 4; Gaps 4;  
QY 3803 AAGATGAGCCCTCCAGAGCTGCAGTCACTCCCTGACCTGGGGCTCATGGCTTCTGACTT 3862  
DB 4 ACGAGGCGAGCCCTCCAGAGCTGCAGTCACTCCCTGACCTGGGGCTCATGGCTTCTGACTT 63  
QY 3863 CTGACCTTCT 3922  
DB 64 CTGACCTTCT 123  
QY 3923 TGTATGCCAACACCTGTTGAGCCAGGAGACAGAAAGCTGTGAGCCTCTGGCCCTTCTCTGG 3982  
DB 124 TGTATGCCAACACCTGTTGAGCCAGGAGACAGAAAGCTGTGAGCCTCTGGCCCTTCTCTGG 183  
QY 3983 ACCGCTGGGCGTGTGATCGGATCAGCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4042  
DB 184 ACCGCTGGGCGTGTGATCGGATCAGCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 243  
QY 4043 AGCTGGGAGGAGGTACAGTAGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4102  
DB 244 AGCTGGGAGGAGGTACAGTAGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 303  
QY 4103 AG-GGAGTGTGAAGTGG-TTCAGGTTGGTGCAGAGCGCTCATGGGCTCC-TGCTTCTTG 4159  
DB 304 AGNGGAGTGTGAAGTGGTTCAGGTTGGTGCAGAGCGCTCATGGGCTCCCTCTCTTNTT 363  
QY 4160 CCTACACTT-GGCCAGTGCACCAGCCCT 4191  
DB 364 GCTACCANTTGGGCAATGCAACCAACAGCCCTT 396

RESULT 12  
US-08-232-463-14  
; Sequence 14, Application US/08232463



Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-9300  
; TELEFAX: (703) 683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 1.9%; Score 99; DB 1; Length 7218;  
Best Local Similarity 5.1%; Pred. No. 9.8e-15;  
Matches 21; Conservative 259; Mismatches 129; Indels 0; Gaps 0;  
Qy 2663 TTGTCCTGTGCGAAGAACTTGTCTTCACTCCACTGCTGCTTCTAGTCCGACCCCTT 2722  
Db 1040 TGGCTGCGAGTCCGAGGAGCTTGCATYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1099  
Qy 2723 TTCTCTCTGCTTCCCTTCCCTGCAATTTCCCAAGAGTGGTCTACACCTCTGCTCC 2782  
Db 1100 YY 1159  
Qy 2783 ACTTCTCTCCACCCACTCACTTCTTAACCCCTTCACTGCTTCCAGGCCCCAGCAA 2842  
Db 1160 YY 1219  
Qy 2843 TGGTCTCTCCAGGTCTGCTAGGACCTCTTCCAGGCCCGACAGTGTGTTGAAGGCTC 2902  
Db 1220 YY 1279  
Qy 2903 ATTCTCTGTGCTGTGTTTGGAGCCACACTGTGTGAGCGTGTGCTTCTCGACTCC 2962  
Db 1280 YY 1339  
Qy 2963 TCTTCTGTGCTGTGCTCTCTGCGCCACTCTTCACTCTCTCACTCTCTCCAGGCTC 3022  
Db 1340 YY 1399

Qy 3023 CTCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3071  
Db 1400 YY 1448

RESULT 13  
US-09-249-585A-4  
; Sequence 4, Application US/09249585A  
; Patent No. 6417002  
; GENERAL INFORMATION:  
; APPLICANT: Horlick, Robert  
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISODES  
; FILE REFERENCE: 0867/0D905  
; CURRENT APPLICATION NUMBER: US/09/249,585A  
; CURRENT FILING DATE: 1999-02-11  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patent In version 3.0  
; SEQ ID NO 4  
; LENGTH: 1926  
; TYPE: DNA  
; ORGANISM: Epstein Barr Virus  
; FEATURES:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1926)  
; OTHER INFORMATION: template strand of EBNA-1 DNA  
US-09-249-585A-4

Query Match 1.2%; Score 60.2; DB 4; Length 1926;  
Best Local Similarity 46.6%; Pred. No. 2.4e-05;  
Matches 335; Conservative 0; Mismatches 373; Indels 11; Gaps 4;  
Qy 2461 ATGCTGTCTTCCATAGCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520  
Db 41 ATCTCTCTTCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100  
Qy 2521 CTGCAACCTGCTCTCTATCAGCTGAACCTTCTTTCGGAGTGTAGTGAGTACCCGTC-- 2578  
Db 101 CTCCCCCACTATTGGTACTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 160  
Qy 2579 -TCTCCAGAGCCCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2637  
Db 161 GTCTCTGGGGGCGCGGAGTCTCTAGTCCGGTCTGTATCTCTACACAGGCTCTGGGG 220  
Qy 2638 CCAATGGGCCACTCTCATCT 2697  
Db 221 TTTTGGAGGTTCAACGTAAACGAGTTTCCCTGGTGCCACTGTCTCTCTCTCTCTCTCT 280  
Qy 2698 CTGCTCTCTCTAGTTCCTGACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2757  
Db 281 CTGCGCTCTCCGCT 340  
Qy 2758 GGAGTGTCTACACCT 2814  
Db 341 CCGCT 400  
Qy 2815 CTGCAATCTGCTCTTCCAGGCCCGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2874  
Db 401 CT 460  
Qy 2875 GCCAAGCCCGCAGTGTGTTTGAAGGCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2934  
Db 461 GTCT 520  
Qy 2935 GCTCAGCGCTGCTGCT 2994  
Db 521 CCGCT 580  
Qy 2995 CTTCTACTCTCTCAGCTCTCTCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3054  
Db 581 CT 639  
Qy 3055 ACTC---TCCCAAGGTTTGGCCACCCAGCAATCAGCACGCTCTCTCTCTCTCTCTCT 3110

[illegible]

## RESULT 14

```

US-09-130-114-2
;
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episodes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

```

Query Match	1.2%	Score 60.2	DB 2	Length 1931
Best Local Similarity	46.6%	Prod. No. 2.4e-05		
Matches 335	Conservative	0	Mismatches 373	Indels 11
Gaps				
Qy	2461	ATGCTGTCTTCATAGACGCTTCCTCTGTCCTACTCATGAGACTGCTCCATTTCTTCTT	2520	
Db	41	ATCTCTCTCTCCCTCTGTGTAGACCTGGTCTTCGAGGCGCGGTACCTGGAGTTCCTT	100	
Qy	2521	CTGCAACCTGTCTCTATACAGCTGAACCTTCTTTCGGAGTGTAGTGAGTACCCGTC--	2578	
Db	101	CTCCCCCACTATTGGTACCTGCTCTGCCCCCTTCTCTGCTCTCTGCTCTCTCGCCCTTCG	160	
Qy	2579	-TCTCCCAAGCCCTCAGCTGGTGGCCCTGGGTGTGACGGCAAAATGGGGCTCTGGTT	2637	
Db	161	GTCTCGGGCGCGCGAGTCTAGTCCCGGTCTGTATCTATCCACAGGCTCTCGGG	220	
Qy	2638	CCAATGGGGCACTCTCATCTCTCTTGTCTTGTGGAGAGAAA	2697	
Db	221	TTTTTCAGGTTCAACGTAAACGACGTTTCCCTGGGTGCCACTTGTCTCTCTCTCGTC	280	
Qy	2698	CTGCCCTCTATAGTCCCGACCCCTTTTCTCTCTGGCTTTCTCTGCCAAAATTTCTCCAA	2757	
Db	281	CTCGCCCTCCCGTCTCTGCTCTCCCGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTC	340	
Qy	2758	GGAGTGGTCTACACCT--CTGCCCTCACTTCTCTCCACCCACATCTCTTTAACCC	2814	
Db	341	CCGCTCTCTCCCGTCTCTCCCGTCTCTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTC	400	
Qy	2815	CTGCAATCTGGCTTCAGGCCCCAGCAATGGTTCCTCCAAAGTCTCTCAGGACATCTT	2874	
Db	401	CTCCCGTCTCTCCCGTCTCTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC	460	
Qy	2875	GCCAAGCCGACAGTGTTTTGAAGGTCATTTCTCTGTGTCTGTCTGTCTGTCTGTCTGT	2934	
Db	461	GTCTCTCCCGTCTCTGTCCTCTCTCCCGTCTCTCCCGTCTCTCTCTCTCTCTCTCTCTC	520	
Qy	2935	GTGAGGGCTGTCCTTCTCGAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2994	
Db	521	CCCGTCTCTGTCCTC	580	
Qy	2995	CTTCTACTCTCCAGCTCTCCAGGCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	3054	
Db	581	CTCCCGTCTCTGTCCTC	639	

Qy	3055	ACTC- ---TCCCAAGTTTGCCACCCAGCCAAATCAGACACGTCCTCTCTCTGACGCGCTTGT	3111
Db	640	CCTCGTCCTCCCGTCCTCGTCTCTCTCTCCCGTCCTCCCGTCCTCCGTCCTCGTCTCT	699
Qy	3111	GCGTCTCTCTCTCTCTCTCTTTCTTAGCGCTCTCAATTGGAGAGCTACACAGCCACACTG	3169
Db	700	CCCGCTCTCGTCTCTCCCGTCTCGTCTCTCCCGTCCTGCTCTCCCGTCCTGCTCTCCCGCTCTCCCG	758

## RESULT 15

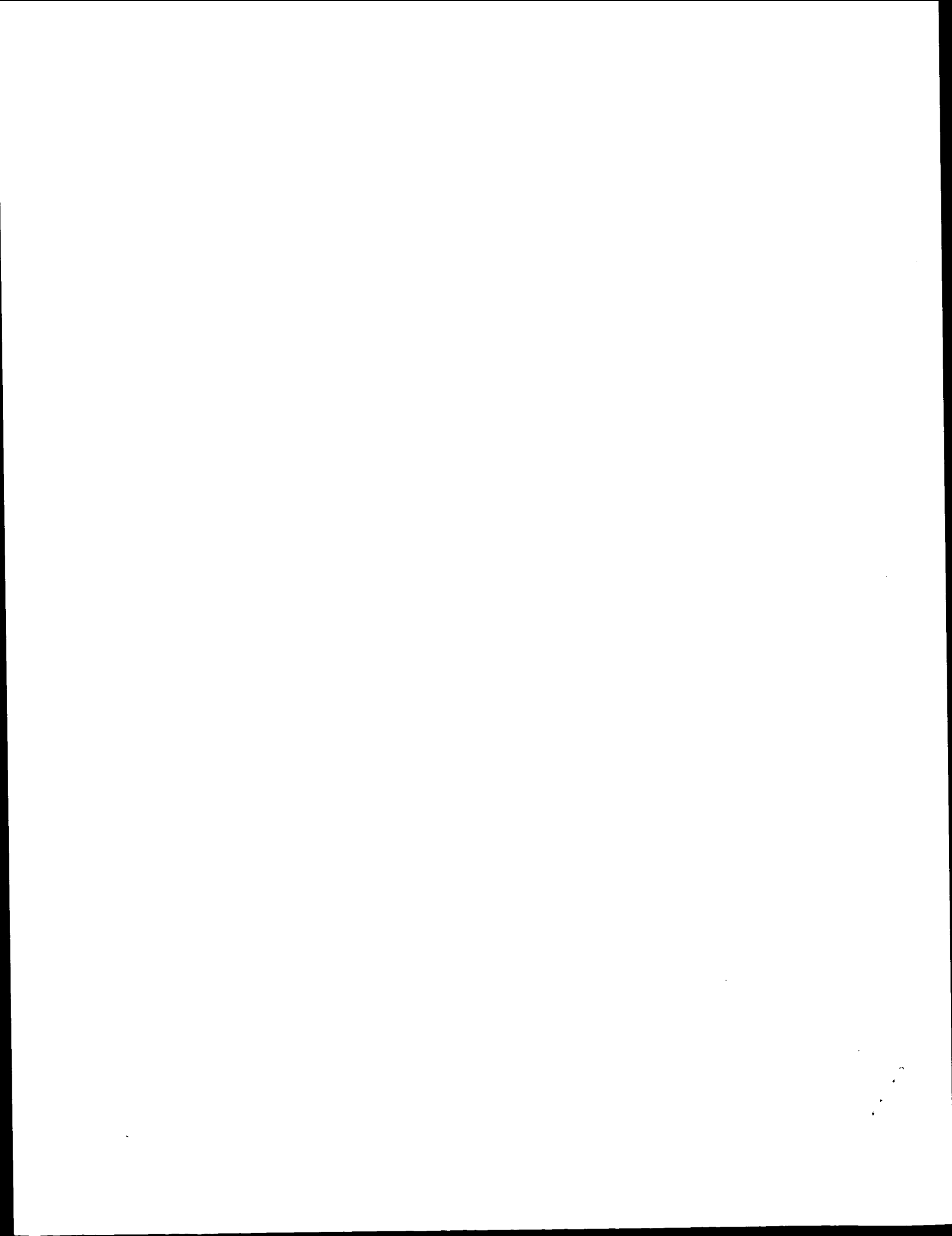
```

US-08-658-136-2
; Sequence 2, Application US/08658136
; Patent No. 607171
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5326 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-658-136-2

```

	Query Match	1.1%;	Score 57.6;	DB 3;	Length 53526;
	Best Local Similarity	45.4%;	Pred. No. 0.00047;		
	Matches 367;	Conservative	0;	Mismatches 429;	Indels 12; Gaps
QY	2319	CCTCCCGCTGGCGCCTTGCCCTATCCAGAGCCTCTCCCTTGTCTTCCCTTCTTCCAG	2378		
DB	34293	CCTCTCCCTCCCGCTAGCGCTTTCCTCTTCTCCCGCAGCGCTTCCCTCTCTCCGCTC	34352		
QY	2379	TTATACATCTCCCTCATCCCTTTCCTGGGCCCCAGCGCTCCCCGAGGTTGGAAAGG	2438		
DB	34353	CCCTAGCGCTTCCCTCTCTCCCTCCCTTACCGCTTCCCTCTCTCC--CCTCCCGCTAGAC	34410		
QY	2439	GCTCTGCGCCTCTTCCCTATACCATGCTGTCTCCAFAGCGCTTCTCTCTGCTTCTACTG	2498		
DB	34411	CTTTCGCTCAGCTCTCCGCTGAGCGCCCTCCACTCGTCCCCAGCGCCCTCCCTCCCGCTA	34470		
QY	2499	AGACTGCGCTCCATTTCTTCTCTGTGCAACCCCTGCTCTCTATCAGCTGAAACCCCTTCTTCGG	2558		

Search completed: February 1, 2003, 08:50:00  
Job time : 394.925 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 2017.75 Seconds  
(without alignments)  
11812.771 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 819  
Sequence: 1 tattattaaagcttcgcgg.....agggccagtcagctgtgctt 819

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	819	100.0	819	6	AR002597	Sequence
2	819	100.0	819	6	AX392089	Sequence
3	819	100.0	819	10	AF176913	Mus muscu
4	674	82.3	881	6	AX205024	Sequence
5	674	82.3	881	6	AX205042	Sequence
6	669.4	81.7	797	6	AR002595	Sequence
7	669.4	81.7	797	6	AX392086	Sequence
8	669.4	81.7	797	9	AF176911	Sequence
9	654	79.9	1736	9	BC012939	Homo sapi
10	627	76.6	1689	9	AF172854	Homo sapi
11	590.4	72.1	680	9	AY049779	Homo sapi
12	540.8	66.0	269355	2	AC109138	Mus muscu
13	509.2	62.2	1692	6	AX205060	Sequence
14	429.6	52.5	5087	6	AR002596	Sequence
15	429.6	52.5	5087	6	AX392088	Sequence
16	429.6	52.5	5087	9	AF176912	Homo sapi
17	429.6	52.5	135116	2	AP002437	Homo sapi
18	429.6	52.5	168567	9	AP003419	Homo sapi
19	429.6	52.3	169144	2	AC005849	Homo sapi
20	236	28.8	283	6	AX202145	Sequence
21	137.6	16.8	63347	2	AC110526	Mus muscu
22	50.8	6.2	125020	9	AF429315	Homo sapi
23	46.2	5.6	47958	2	AC091104	Homo sapi
24	45.4	5.5	32134	2	AC096823	Rattus no
25	45.2	5.5	125020	9	AF429315	Homo sapi
26	44.4	5.4	187727	2	AC021142	Homo sapi
27	43.2	5.3	6803	9	HSU86758	Human netri
28	42.4	5.2	1357	9	AF035771	Homo sapi
29	42.4	5.2	1524	9	AF004900	Homo sapi
30	42.4	5.2	1578	6	AX410794	Sequence
31	42.4	5.2	1578	9	HSU82108	Sequence
32	42.4	5.2	1600	9	HSTKALMR	Human SIP-1
33	42.4	5.2	1639	9	BC014513	ZS0150 H.sapiens m
34	42.4	5.2	1642	6	AR070449	Sequence
35	42.2	5.2	1901	5	GGLAMB1	Chicken mRN
36	42	5.1	63082	2	AC022663	Homo sapi
37	41.2	5.0	115666	2	AC105744	Oryza sat
38	41.2	5.0	119495	2	AC126302	Rattus no
39	41.2	5.0	123071	2	AC121481	Rattus no
40	41.2	5.0	173322	2	AC127095	Rattus no
41	41	5.0	158537	2	AC126887	Rattus no
42	40.8	5.0	1983	9	HSM803625	Homo sapi
43	40.8	5.0	2750	9	AK097089	Homo sapi
44	40.8	5.0	4015	9	AB075829	Homo sapi
45	40.8	5.0	71418	9	AL732442	Human DNA

ALIGNMENTS

RESULT 1  
AR002597  
LOCUS  
DEFINITION Sequence 4 from patent US 5741772.  
ACCESSION AR002597  
VERSION AR002597.1 GI:3964151  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 819)  
AUTHORS Chang, M.-s.  
TITLE Neurotrophic factor NNT-1  
JOURNAL Patent: US 5741772-A 4 21-APR-1998;  
FEATURES Location/Qualifiers

819 bp DNA linear PAT 04-DEC-1998



```

QY 601 GGGCCCTGGCCCTGGCCACAGTGACTTCTCCAGAGATGGATGACTTCTGGCTGCTGAA 660
Db 601 GGGCCCTGGCCCTGGCCACAGTGACTTCTCCAGAGATGGATGACTTCTGGCTGCTGAA 660
QY 661 GGAGCTGCAGACCTGGCTATGCGCTTCCAGCAAGAGATTCACACCGGCTTAAGAAGAT 720
Db 661 GGAGCTGCAGACCTGGCTATGCGCTTCCAGCAAGAGATTCACACCGGCTTAAGAAGAT 720
QY 721 GCAGCTCCAGCAGCTTCAGTCAACCTCGACCTTGAGGACACATGGTTTCTGACCTCTGAC 780
Db 721 GCAGCTCCAGCAGCTTCAGTCAACCTCGACCTTGAGGACACATGGTTTCTGACCTCTGAC 780
QY 781 CTTTAAACCCACACTCCAGGCCAGTCAGCTGTGCTT 819
Db 781 CTTTAAACCCACACTCCAGGCCAGTCAGCTGTGCTT 819

RESULT 3
AF176913 819 bp mRNA linear ROD 04-OCT-1999
LOCUS Mus musculus neurotrophin-1/B-cell stimulating factor-3 mRNA,
DEFINITION complete cds.
ACCESSION AF176913
VERSION AF176913.1 GI:6007644
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 819);
Senaldi, G., Varnum, B.C., Sarmiento, U., Starnes, C., Lile, J.,
Scully, S., Guo, J., Elliott, G., McNinch, J., Shaklee, C.L.,
Freeman, D., Manu, F., Simonet, W.S., Boone, T. and Chang, M.-S.
TITLE Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the
IL-6 family
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)
MEDLINE 99432254
PubMed 10500198
REFERENCE 2 (bases 1 to 819)
AUTHORS Senaldi, G., Varnum, B., Sarmiento, U., Lile, J., Starnes, C.,
Scully, S., Guo, J., Elliott, G., McNinch, J., Freeman, D.,
Manu, F., Simonet, S., Boone, T. and Chang, M.-S.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1999) Amgen, Inc., One Amgen Center Drive,
Thousand Oaks, CA 91320, USA
FEATURES
source Location/Qualifiers
1..819
/organism="Mus musculus"
/db_xref="taxon:10090"
95..772
/codon_start=1
/product="neurotrophin-1/B-cell stimulating factor-3"
/protein_id="AAP00993.1"
/db_xref="GI:6007645"
/translation="MDLRAGSGWGLACLTVMHLPAVPAALNRGDPGPGSIQKTY
DLTRYELHQLRSAGTYLYNLPFPNPPRLGAEITPRATVNVWVRSNDRLR
LTQNEYASHLLCYLRGLNQAATFALRSLAHFCTSLQGLLGSIAQWATLGVLPQ
PLPCTEPAPWAPGSHSDFLQKMDDFLLKELQTLWLSAKDFNRLKKMQPPAASVTL
HLEAHGF"
BASE COUNT 156 a 288 c 218 g 157 t
ORIGIN
Query Match 100.0%; Score 819; DB 10; Length 819;
Best Local Similarity 100.0%; Pred. No. 3.4e-176;
Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCATCCGACAGCTCTGGGAGA 60
Db 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCATCCGACAGCTCTGGGAGA 60
QY 61 GGAGCCGGCCGGCCGGCCGGCCGGCCGGCCGGCCGGCCGGCCGGCCGGCCGGCCGGCCGG 120

```

```
CDS
174..851
/note="NNT-1"
/codon_start=1
/protein_id="CAC60175.1"
/db_xref="GI:15394260"
/translation="MDLRAGDSWGLMCLCTVLMHLPVAPALNRTGDPGPGSIQKTY
DLTRYLEHQLRSLAGTYLNYLGPFPNPPRLGAETLPRATVDLEVRSLNDKLR
LTQNYEAYSHLLCYLRGLNRQAATAELRSLAHFCTSLQGLLGSITAGVMAALGYPLPQ
PLPTEPTWTPGPAHSDFLQMDDFLLKELQTLWLRSAKDFNRLKKKQPPAAAVTL
HLGAHGF"
BASE COUNT 158 a 318 c 246 g 159 t
ORIGIN
Query Match 82.3%; Score 674; DB 6; Length 881;
Best Local Similarity 92.2%; Pred. No. 3.5e-143; Indels 1; Gaps 1;
Matches 721; Conservative 0; Mismatches 60;

QY 2 ATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGACGCTCTGGGAGAG 61
DB 82 ATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGACGCTCTGGGAGAG 141
QY 62 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 121
DB 142 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 200
QY 122 GGGATGTTAGCTTGCTGATGCAAGTGTCTGGACCTCCCTCCCACTCCGACGCTCTTAAT 181
DB 201 GGGATGTTAGCTTGCTGATGCAAGTGTCTGGACCTCCCTCCCACTCCGACGCTCTTAAT 260
QY 182 CGCACAGGAGATCCAGGCGCTCGCGGCTCGCCCTCCCACTCCGACGCTCTTAAT 241
DB 261 CGCACAGGAGATCCAGGCGCTCGCGGCTCGCCCTCCCACTCCGACGCTCTTAAT 320
QY 242 CTGGAGCATCAACTCCGAGCTTAGCTGGAGCTTACCTGAACTACCTGAGGCGCCCTTTC 301
DB 321 CTGGAGCATCAACTCCGAGCTTAGCTGGAGCTTACCTGAACTACCTGAGGCGCCCTTTC 380
QY 302 AACGAGCTGACTTCAATCTCTGCTGACGCTGGGCGAGAACTCTGCGCGCGCGCGCGCTC 361
DB 381 AACGAGCTGACTTCAATCTCTGCTGACGCTGGGCGAGAACTCTGCGCGCGCGCGCGCTC 440
QY 362 AACTTGAAGTGTGGCGAAGCTTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGCG 421
DB 441 GACTTGAAGTGTGGCGAAGCTTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGCG 500
QY 422 TACAGTCACTCTCTGTTACTTGGTGGCTTCAACCGTCAAGTGTGCGAGCTGAGGCTGAG 481
DB 501 TACAGTCACTCTCTGTTACTTGGTGGCTTCAACCGTCAAGTGTGCGAGCTGAGGCTGAG 560
QY 482 CGACGTAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 541
DB 561 CGACGTAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 620
QY 542 GTCATGGCGAGCTTGGCTACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 601
DB 621 GTCATGGCGAGCTTGGCTACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680
QY 602 GCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 661
DB 681 ACTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 740
QY 662 GAGCTGAGACCTTGGCTATGCGGTTACGCAAGGACTTCAACCGGCTTAAAGAAAGATG 721
DB 741 GAGCTGAGACCTTGGCTATGCGGTTACGCAAGGACTTCAACCGGCTTAAAGAAAGATG 800
QY 722 CAGCCTCCAGCAGCTTCACTCAGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
DB 801 CAGCCTCCAGCAGCTTCACTCAGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCG 860
QY 782 CT 783
DB 861 TT 862
```





Thousand Oaks, CA 91320, USA	
FEATURES	Location/Qualifiers
source	1. .797
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
CDS	90...767
	/codon_start=1
	/product="neurotrophin-1/B-cell stimulating factor-3"
	/protein_id="AAF00991.1"
	/db_xref="GI:6007641"
	/translation="MDLRAGDSWGMCLCLTVLWHLPAVPAVLRNRTGDPGPGPSIQITY
	DTNRYLEHQIRSLACTYLNLYGPPNEPDPNPRICAEATLPRAIVLEWVRSLGVPLPQ
	LTYQYEAHSLHCLYLRGNRQATAEIRSLAHPCTSLQGLGSIAGVMAALGVPLPQ
	PLFGTEPTWTPGPAHSDFLQKMDDFLLWLKELQTLWLRSAKDFNRLKKMQPPAAVATL
	HLGAGHF"
BASE COUNT	139 a 297 c 218 g 143 t
ORIGIN	
	Query Match 81.7%; Score 669.4; DB 9; Length 797;
	Best Local Similarity 92.0%; Pred. No. 3.9e-142;
	Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
QY	5 ATTHAAAGCTTTGCGCGAGCGCGGCTCGCCCTCCACTCCGCGACGCTCTCGGAGAGGAG 64
DB	1 ATTHAAAGCTTTGCGCGAGCGCGGCTCGCCCTCCACTCCGCGACGCTCTCGGAGAGGAG 60
QY	65 CCGCGCCCG 124
DB	61 CCGCACCCCG 119
QY	125 ATGTTAGCTTGCTTATGCA CGGTGCTGTGGCACCTCCCTGCGAGTGCAGCTCTTAATCGC 184
DB	120 ATGTTAGCTTGCTTATGCA CGGTGCTGTGGCACCTCCCTGCGAGTGCAGCTCTCAATCGC 179
QY	185 ACAGAGATCCAGGCGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 244
DB	180 ACAGGGGACCCAGGCG 239
QY	245 GAGCATCAACTCCGCGAGCTTAGCTGGGACCTACCTGAACTACCTCTGGGCGCGCGCGCG 304
DB	240 GAGCACCAACTCCGCGAGCTTAGCTGGGACCTATCTGAACTACCTCTGGGCGCGCGCG 299
QY	305 GAGCTGATCTCAATCTCTCGACTGGGGCGAGAACTCTGCCCGAGGCGCGCGCGCGCG 364
DB	300 GAGCGAGACTTCAACCCCTCCCGCTGGGGCGAGAGACTCTGCGCGAGGCGCGCGCGCG 359
QY	365 TTGGNAGCTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGCGGTAC 424
DB	360 TTGGAGGCTGGCGAAGCCTCAATGACAACTGCGGCTGACCCAGAACTATGAGGCGGTAC 419
QY	425 AGTCACCTCTCTGTGTACTTTGCTGGGCGCTCAACCGTCAGGCTGCGACAGCTGAACTCCGA 484
DB	420 AGCCACCTCTCTGTGTACTTTGCTGGGCGCTCAACCGTCAGGCTGCGACCTGCTGAGCTGCGC 479
QY	485 CGTAGCTGGGCGCACTTCTGTATACAGCTCTCAGGGCGCTGCTGGGCGAGCAATTCAGGTGTC 544
DB	480 CGCAGCTGGGCGCACTTCTGTATACAGCTCTCAGGGCGCTGCTGGGCGAGCAATTCAGGTGTC 539
QY	545 ATGGCGAGCGCTGGCTTACCCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 604
DB	540 ATGGGAGCTGCTGGGCTTACCCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 599
QY	605 CTGCGCGCTGCGCGACAGTGACTTCTCCAGAAAGATGGATGACTTCTGGCTGTGAAAGGAG 664
DB	600 CTGGCGCTGCGCGACAGTGACTTCTCCAGAAAGATGGATGACTTCTGGCTGTGAAAGGAG 659
QY	665 CTGCGAGACTGGCTATGGCGTTACGCCAAGGACTTCAACCGGCTTAAGAGAGATGCGAG 724
DB	660 CTGCGAGACTGGCTATGGCGTTACGCCAAGGACTTCAACCGGCTTAAGAGAGATGCGAG 719
QY	725 CCTCCAGAGCTTCAGTCAACCTCTGAGGAGCAGATGGTTTCTGACCTCTGACCTT 783
DB	720 CCTCCAGAGCTTCAGTCAACCTCTGAGGAGCAGATGGTTTCTGACCTCTGACCTT 778
RESULT 8	
AF176911	
LOCUS	797 bp mRNA linear PRI 04-OCT-1999
DEFINITION	Homo sapiens neurotrophin-1/B-cell stimulating factor-3 mRNA,
complete cds.	
ACCESSION	AF176911
VERSION	AF176911.1 GI:6007640
KEYWORDS	.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
	1 (bases 1 to 797)
	Senaldi,G., Varnum,B.C., Sarmiento,U., Starnes,C., Lile,J.,
	Scully,S., Guo,J., Elliott,G., McNinch,J., Shaklee,C.I.,
	Freeman,D., Manu,F., Simonet,W.S., Boone,T. and Chang,M.S.
	Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the
	IL-6 family
	Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)
	99432254
	10500198
	2 (bases 1 to 797)
	Senaldi,G., Varnum,B., Sarmiento,U., Lile,J., Starnes,C.,
	Scully,S., Guo,J., Elliott,G., McNinch,J., Freeman,D., Shaklee,C.,
	Manu,F., Simonet,S., Boone,T. and Chang,M.-S.
	Direct Submission
	Submitted (11-AUG-1999) Amgen, Inc., One Amgen Center Drive,
	JOURNAL
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
TITLE	

```

RESULT 9
BC012939
LOCUS
DEFINITION
Homo sapiens, similar to cardiotrophin-like cytokine;
neurotrophin-1/B-cell stimulating factor-3, clone MGC:21195
IMAGE:4453813, mRNA, complete cds.
ACCESSION
BC012939
VERSION
BC012939.1 GI:15277894
KEYWORDS
MGC.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1736)
Direct Submission
Submitted (20-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalobos@bcm.tmc.edu.
Villalobos, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAC Plate: 28 Row: b Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6007640.
FEATURES
source
1..1736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:21195 IMAGE:4453813"
/tissue type="Kidney, hypernephroma"
/clone lib="NIH_MGC_89"
/lab host="DH10B"
/note="vector: pCMV-SPORT6"
73..750
/codon start=1
/product="Similar to cardiotrophin-like cytokine;
neurotrophin-1/B-cell stimulating factor-3"
/protein id="AAH12939.1"
/db_xref="GI:15277895"
/translation="MDLRAGDSWGMGLACTVLVHLPAVLPALNRTPDGPSPGIQKTY
DLTRYLEHSLAGTYLNYLPGFPNPFPLGAEITLPRATVLEWVRSINDKLK
LTQNYEAYSHLLCYLGNLRQAETAEIRLSIAHFTSLQGLIGSIAGVMAALGVPLPQ
PLPGETPTWTPGPAHSDFLQKMDDFLLKELQTLWRSKDFNRLKKMQPFAAVTL
HLGAGGF"
BASE COUNT 371 a 546 c 454 g 365 t
ORIGIN
Query Match 79.9%; Score 654; DB 9; Length 1736;
Best Local Similarity 92.0%; Pred. No. 1.1e-138;
Matches 701; Conservative 0; Mismatches 60; Indels 1; Gaps 1;
QY 22 GCGCGGCTCGCCCTCCCACTCCGCCAGCTCTGGGAGAGAGCGCGCGCGCGCC 81
Db 1 GCGCGGCTCGCCCTCCCACTCCGCCAGCTCTGGGAGAGAGCGCGCGCGCGCC 60

```

```

QY 82 GGCCCCCAGCCCCCATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCTTGGCCTATG 141
Db 61 GG-CCCAAGCCCCCATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCTTGGCCTATG 119
QY 142 CACGGTGTGTGGACCTCCCTGCACTGCGAGTCTTAATCGCACAGGAGATCCAGGCC 201
Db 120 CACGGTGTGTGGACCTCCCTGCACTGCGAGTCTTAATCGCACAGGAGATCCAGGCC 179
QY 202 TGGCCCTCCATCCAGAAAACCTATGACCTCACCCCTACCTGGAGATCAATCCGCGAG 261
Db 180 TGGCCCTCCATCCAGAAAACCTATGACCTCACCCCTACCTGGAGATCAATCCGCGAG 239
QY 262 CTTAGCTGGGACCTACTCTGAACTTACCTGGGGCCCCCTTTCAACAGACCTGACTTCAATCC 321
Db 240 CTTAGCTGGGACCTACTCTGAACTTACCTGGGGCCCCCTTTCAACAGACCTGACTTCAATCC 299
QY 322 TCTTGACCTGGGGGCGAGAACTCTGCCCAGGGCCACGGTCAACTTTGGAAAGTGTGGCGAAG 381
Db 300 TCCCCGCTGGGGGCGAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAG 359
QY 382 CCTCAATGACAGGCTGGCGCTGACCCAGAACTATGAGGGCTACAGTCACTCTCTGTGTTA 441
Db 360 CCTCAATGACAACTGGCGCTGACCCAGAACTATGAGGGCTACAGTCACTCTCTGTGTTA 419
QY 442 CTTGCTGGCCCTCAACCGTCAAGCTGCCACAGCTGAACTCCGACGTAGCTTGGCCCACTT 501
Db 420 CTTGCTGGCCCTCAACCGTCAAGCTGCCACAGCTGAACTCCGACGTAGCTTGGCCCACTT 479
QY 502 CTGTACCAAGCTTCCAGGGCTCTGCGGAGCATTCGAGGTGTCTATGGCGACGCTTGGCTA 561
Db 480 CTGCACCAAGCTTCCAGGGCTCTGCGGAGCATTCGCGGCGTTCATGGCGAGCTCTGGGCTA 539
QY 562 CCACCTGCCCCAGGCTCTGCGGAGCATTCGAGCCAGCTGAGCCCTGGGCCCTTGGCCCTGCCACAG 621
Db 540 CCACCTGCCCCAGGCTCTGCGGAGCATTCGAGCCAGCTGAGCCCTGGGCCCTTGGCCCTGCCACAG 599
QY 622 TGACTTCTCCAGAGATGGATGACTTCTGCTGCTGAAGAGAGCTGCAGACCTGGCTATG 681
Db 600 TGACTTCTCCAGAGATGGATGACTTCTGCTGCTGAAGAGAGCTGCAGACCTGGCTATG 659
QY 682 GCGTTCAGCCAGGACTTCAACCGCTTAAAGAGAGATGAGCCCTCCAGCAGCTTCACT 741
Db 660 GCGTTCAGCCAGGACTTCAACCGCTTAAAGAGAGATGAGCCCTCCAGCAGCTTCACT 719
QY 742 CACCTGCACTTGGAGGCACATGTTTCTGACCTCTGACCTT 783
Db 720 CACCTGCACTTGGAGGCATGTTTCTGACCTTCTGACCTT 761
RESULT 10
AF172854
LOCUS
DEFINITION
Homo sapiens cardiotrophin-like cytokine CLC mRNA, complete cds.
ACCESSION
AF172854
VERSION
AF172854.1 GI:5852980
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1689)
Ruben, S. and Alderson, R.F.
Shi, Y., Wang, W., Yourey, P.A., Gohari, S., Zukauskas, D., Zhang, J.,
Computational EST database analysis identifies a novel member of
the neuropoietic cytokine family
Biochem. Biophys. Res. Commun. 262 (1), 132-138 (1999)
99382254
PUBMED
10448081
REFERENCE
2 (bases 1 to 1689)
AUTHORS
Shi, Y.
Direct Submission
Submitted (28-JUL-1999) Molecular Biology, Human Genome Science,
Inc., 9410 Key West Avenue, Rockville, MD 20850, USA

```

Db 720 CTGACTTCTGACCTT 734

RESULT 11  
AY049779 680 bp mRNA linear PRI 23-OCT-2001  
LOCUS Homo sapiens cardiotrophin-like cytokine (CLC) mRNA, complete cds.  
DEFINITION  
ACCESSION AY049779  
VERSION AY049779.1 GI:16356642  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Hu, X., Xu, Y., Zhang, B., Peng, X., Yuan, J. and Qiang, B.  
TITLE Direct Submission  
JOURNAL Submitted (30-JUL-2001) Department of Biochemistry, Institute of Basic Medical Science, Chinese Academy of Medical Sciences, 5 Dong Dan San Tiao, Beijing, 100005, P.R. China  
FEATURES  
source Location/Qualifiers  
1..680  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..680  
/gene="CLC"  
3..680  
/gene="CLC"  
/note="neurotrophin-1/B-cell stimulating factor-3"  
/codon\_start=1  
/product="cardiotrophin-like cytokine"  
/protein\_id="AA115436.1"  
/db\_xref="GI:16356643"  
/translation="MDLRAGDSWGMCLACTVLVHLPVPAALNRGDPGPGPSIOKTY  
DLTRYLEHQLRSAGTYLVNLPFPNPFRLGAETLPRAIVDLEWVRSINDKLRL  
LTQNYEAYSHLLCYLRGLNQATAEIRSLAHFCTSLQGLLGSAGVMAALGYPLPQ  
PLPGTEFTWTPGPAHSDFLQKMDDFLLKELQTLWRSKADFNRLKKKQPPAAAVTL  
HLGAHGF"

BASE COUNT 125 a 239 c 191 g 125 t

ORIGIN

Query Match 72.1%; Score 590.4; DB 9; Length 680;  
Best Local Similarity 91.8%; Pred. No. 3.9e-124;  
Matches 624; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 93 CCATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTTGCTATGCACGGTGTCT 152  
Db 1 CCATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTTGCTATGCACGGTGTCT 60

Qy 153 GGCACTTCCCTGCACTGCCAGCTCTTAATCGCAGGAGATCCAGGCCCTGGCCCCCTCCA 212  
Db 61 GGCACTTCCCTGCACTGCCAGCTCTTAATCGCAGGAGATCCAGGCCCTGGCCCCCTCCA 120

Qy 213 TCCAGAAAACCTATGACTACACCGCTACTTGGAGCATCACTCCGAGCTTAGCTGGGA 272  
Db 121 TCCAGAAAACCTATGACTACACCGCTACTTGGAGCATCACTCCGAGCTTAGCTGGGA 180

Qy 273 CCTACCTGAACCTACTCGGGCCCCCTTTCAACAGAGCTGACTTCAATCCTCTCGACTGG 332  
Db 181 CCTATCTGAACCTACTCGGGCCCCCTTTCAACAGAGCTGACTTCAACCTCTCCGCTGG 240

Qy 333 GGGCAGAACTCTGCCAGGCCACCGGTCAAATTGGGAAGTGTGGGGAAGCCCTCAATGACA 392  
Db 241 GGGCAGAGACTCTGCCAGGCCACCTGTGACTTGGAGGTGTGGGGAAGCCCTCAATGACA 300

Qy 393 GGCTCGGCTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTGTACTTTCGTGGGCC 452  
Db 301 AACTCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACTTTCGTGGGCC 360

Qy 453 TCAACCGCTCAGCTGCCACAGCTGAACTCCGACGTAGCTGGCCACCTTCTGTATACAGCC 512  
Db 361 TCAACCGCTCAGCTGCCACCTGCTGAGCTGCGCGCAGCTGGCCACCTTCTGTATACAGCC 420

QY 513 TCAGGCGCTGCTGGGAGGATTCAGGTGTCATGGCGAGCGCTTGGCTACCCACTGCCCC 572  
 Db 421 TCCAGGCGCTGCTGGGAGGATTCAGGTGTCATGGCGAGCGCTTGGCTACCCACTGCCCC 480  
 QY 573 AGCCTCTGCCAGGAGTGCAGCGCTGGGCGCTGCGCCCTGCGCCAGTGCAGTCTCTCTCC 632  
 Db 481 AGCGCTGCTGGAGTGAACCACTTGGACTCTCTGCGCCCTGCGCCAGTGCAGTCTCTCTCC 540  
 QY 633 AGAAGATGGATGACTTCTGCTGCTGAAGGAGTGCAGAGCTGCGCTATGGCTTCAGCCA 692  
 Db 541 AGAAGATGGAGACTTCTGCTGCTGAAGGAGTGCAGAGCTGCGCTATGGCTTCAGCCA 600  
 QY 693 AGACTTCAACCGCTTAAGAAGATGCAGCTTCAGAGCTTCAGTCACTTCGACT 752  
 Db 601 AGGACTTCAACCGCTCAAGAAGATGCAGCTTCAGAGCTTCAGTCACTTCGACT 660  
 QY 753 TGGAGGACATGTTCTCTGA 772  
 Db 661 TGGGGCTCATGGCTTCTGA 680

RESULT 12  
 AC109138  
 LOCUS  
 DEFINITION Mus musculus clone RP23-41B18, WORKING DRAFT SEQUENCE, 7 ordered pieces.  
 AC109138  
 VERSION AC109138.4 GI:22381069  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Birren,B., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP23-41B18  
 Unpublished  
 2 (bases 1 to 269355)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Farrow,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Rettar,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 269355)  
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Farrow,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,

McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Rettar,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 21, 2002 this sequence version replaced gi:20451120.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L14003  
 Center clone name: 41\_B\_18  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 267668 bases at least Q40  
 Consensus quality: 268459 bases at least Q30  
 Consensus quality: 268627 bases at least Q20  
 Insert size: 242000; agarose-fp  
 Quality coverage: 268755; sum-of-contigs  
 Quality coverage: 10.4 in Q20 bases; agarose-fp  
 Quality coverage: 9.4 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and the accession number will be preserved.  
 \* 1 100642: contig of 100642 bp in length  
 \* 100643 100742: gap of 100 bp  
 \* 100743 104988: contig of 4246 bp in length  
 \* 104989 105088: gap of 100 bp  
 \* 105089 112582: contig of 7494 bp in length  
 \* 112583 112682: gap of 100 bp  
 \* 112683 142041: contig of 29359 bp in length  
 \* 142042 142141: gap of 100 bp  
 \* 142142 179941: contig of 37800 bp in length  
 \* 179942 180041: gap of 100 bp  
 \* 180042 259317: contig of 79276 bp in length  
 \* 259318 259417: gap of 100 bp  
 \* 259418 269355: contig of 9938 bp in length.  
 FEATURES  
 source  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-41B18"  
 /clone\_lib="RPCI-23 Female Mouse BAC"  
 misc\_feature  
 1 100642  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"  
 misc\_feature  
 100743 .104988  
 /note="assembly\_fragment"  
 misc\_feature  
 105089 .112582  
 /note="assembly\_fragment"  
 misc\_feature  
 112683 .142041

```
/note="assembly_fragment"
142142..179941
/note="assembly_fragment"
180042..259317
/note="assembly_fragment"
259418..269355
/note="assembly_fragment
clone_end:77
vector_side:right"
BASE COUNT 67479 a 67954 c 68125 g 65191 t 606 others
ORIGIN

Query Match 66.0%; Score 540.8; DB 2; Length 269355;
Best Local Similarity 99.6%; Pred. No. 3.8e-113;
Matches 542; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 276 ACCTGAACCTACCTGGGGCCCCCTTTCAAGAGAGCTGACCTCAATCTCTCGACTGGGG 335
Db 151317 AGCTGAACCTACCTGGGGCCCCCTTTCAAGAGAGCTGACCTCAATCTCTCGACTGGGG 151376

QY 336 CAGAAACTCTGCCAGGGCCAGGTCACCTTGAAGTGTGGCGAAGCCTCAATGACAGGC 395
Db 151377 CAGAAACTCTGCCAGGGCCAGGTCACCTTGAAGTGTGGCGAAGCCTCAATGACAGGC 151436

QY 396 TCGCGCTGACCCAGAACTATGAGCGGTACAGTCACCTCTCTGTGTTACTTTCGCTGGCTCA 455
Db 151437 TCGCGCTGACCCAGAACTATGAGCGGTACAGTCACCTCTCTGTGTTACTTTCGCTGGCTCA 151496

QY 456 ACCGTCAGGCTGCCAGCTGAATCCGAGTAGCTTGGCTACCCACTGCCCTCC 515
Db 151497 ACCGTCAGGCTGCCAGCTGAATCCGAGTAGCTTGGCTACCCACTGCCCTCC 151556

QY 516 AGGGCTGTCTGGGAGCATTTGAGGTGTCTATGCGAGCTTGGCTACCCACTGCCCTCCAGC 575
Db 151557 AGGGCTGTCTGGGAGCATTTGAGGTGTCTATGCGAGCTTGGCTACCCACTGCCCTCCAGC 151616

QY 576 CTCTGCCAGGACTGAGCAGCTTGGGCCCCCTGGCCCTGCCCTCCAGTGAATCTCTCCAGA 635
Db 151617 CTCTGCCAGGACTGAGCAGCTTGGGCCCCCTGGCCCTGCCCTCCAGTGAATCTCTCCAGA 151676

QY 636 AGATGATGACTTCTGGCTGCTGAAGAGCTGAGACCTTGGCTATGGCTTACGCCAAGG 695
Db 151677 AGATGATGACTTCTGGCTGCTGAAGAGCTGAGACCTTGGCTATGGCTTACGCCAAGG 151736

QY 696 ACTTCAACCGGCTTAAGAAAGATGACGCTCCAGCAGCTTCAAGTCAACCTGCACCTTGG 755
Db 151737 ACTTCAACCGGCTTAAGAAAGATGACGCTCCAGCAGCTTCAAGTCAACCTGCACCTTGG 151796

QY 756 AGGCATAGTTTCTGACCTCTGACCTTAAACCCACACCTCCAGGCCCAAGTCACTGT 815
Db 151797 AGGCCATAGTTTCTGACCTCTGACCTTAAACCCACACCTCCAGGCCCAAGTCACTGT 151856

QY 816 GCTT 819
Db 151857 GCTT 151860

RESULT 13
LOCUS AX205060 1692 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 19 from Patent WO0155219.
ACCESSION AX205060
VERSION AX205060.1 GI:15394299
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequences.
1 (bases 1 to 1692)
Elson, G. and Gauchat, J. F.
Scsentrfr/mnt-1 fusion protein
Patent: WO 0155219-A 19 02-AUG-2001;
PIERRE FABRE MEDICAMENT (FR)
FEATURES
* Location/Qualifiers
```

```
source 1..1692
/organism="synthetic construct"
CDS 1..1692
/note="proteine de fusion"
/codon_start=1
/transl_table=11
/protein_id="CAC60181.1"
/db_xref="GI:15394300"
/translation="MAAPVWACCAVLAATAAAVVVQHRHSQEPAPHVQYERLGSVDVTL
PCGTANDAATVTRVNGTDLAPDLNGSLVHLGELHGLXYACFHRDSWHLRHQVL
KRVGLPREVLSCRSNTYKGYCSWHLPTPTYPNTFNTVTLHSGSKIMVCEKDPAL
KHCIRYVHLFTSTIKYKVISYNSALGHNATAITFDBFTTVKDDPPNVPVPSN
PRELEVWOTPSWPDPEFPLKFLRYRPLLDQWQVLSLDTAHTTITDAYACKY
IIQVAAKNEIGTWSWVAHAATPTEPRHLTTEQAQETTTSTSSLAPPPTTKI
CDLEGGSGGGSGGGSLRLNLTGPGPSITQKTYDLTRYLHQUKSLAGTYLNY
LGPFPNPDPPNPRLAGETLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRGLNR
QATATBLRSLAHFTSLQGLGSIAQMAALGYPLPQLPGLTPTWTPGFAHSDFLQ
KMDDFWLLKELQTLWRSKDFNRLKKMQPPAAAVTLHLGAHGFEQKLISEDL"
BASE COUNT 337 a 582 c 470 g 303 t
ORIGIN

Query Match 62.2%; Score 509.2; DB 6; Length 1692;
Best Local Similarity 91.1%; Pred. No. 1.1e-105;
Matches 541; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 176 CTTAATCGCACAGAGATCCAGGCCCTGGCCCTCCATCCAGAAACCTATGACCTCAC 235
Db 1066 CTTAATCGCACAGAGATCCAGGCCCTGGCCCTCCATCCAGAAACCTATGACCTCAC 1125

QY 236 CGCTTACCTGGAGCATCAACTCCGAGCTTAGCTGGGACTTACCTGAACTACCTGGGGCC 295
Db 1126 CGCTTACCTGGAGCATCAACTCCGAGCTTAGCTGGGACTTACCTGAACTACCTGGGGCC 1185

QY 296 CTTTAAAGAGCTGACTTCAATCTCTCGATCTGGGGGAGAAACTCTGCCAGGGCC 355
Db 1186 CTTTAAAGAGCTGACTTCAATCTCTCGATCTGGGGGAGAGACTCTGCCAGGGCC 1245

QY 356 ACGGTCAACTTGAAGTGTGCGAAGCTCAATGACAGCTGCGGCTGACCCAGAACATAT 415
Db 1246 ACTGTTGACTTGAAGTGTGCGAAGCTCAATGACAACTGCGGCTGACCCAGAACATAT 1305

QY 416 GAGCGGTACAGTCACTCTCTGTGTACTTGGTGGCCCTCAACCGCTCAGGCTGCCACAGCT 475
Db 1306 GAGCGGTACAGTCACTCTCTGTGTACTTGGTGGCCCTCAACCGCTCAGGCTGCCACAGCT 1365

QY 476 GAAGTCCGAGTAGCTGGCCACTTCTGTACAGACCTCCAGGGCTCTGCGGAGCATT 535
Db 1366 GAGTCCGCGCAGCTGGCCACTTCTGTACAGACCTCCAGGGCTCTGCGGAGCATT 1425

QY 536 GCAGGTGTCATGGCGAGCTTGGCTTACCCACTGCCAGCCCTTCCAGGGAGCTCAGCCA 595
Db 1426 GCGGGCTGTCATGGCGAGCTTGGGCTACCCACTGCCAGCCCTTCCAGGGAGCTCAGCCA 1485

QY 596 GCCTGGGCCCCCTGGCCCTGCCCCACAGTGAATCTCTCCAGAAAGATGAGTACTTCTGGCTG 655
Db 1486 ACTTGGACTCTGGCCCTGCCCCACAGTGAATCTCTCCAGAAAGATGAGTACTTCTGGCTG 1545

QY 656 CTGAAGAGCTGACAGCTGGCTTATGGCTTACGCAAGGACTTCAACCGGCTTAAGAG 715
Db 1546 CTGAAGAGCTGACAGCTGGCTTATGGCTTACGCAAGGACTTCAACCGGCTTAAGAG 1605

QY 716 AAGATGACAGCTCCAGAGCTTCAAGTCACTGACCTGGAGGGACATGTTTC 769
Db 1606 AAGATGACAGCTCCAGAGCTTCAAGTCACTGACCTGGAGGGCTCATGGCTTC 1659

RESULT 14
LOCUS AR002596
DEFINITION Sequence 3 from patent US 5741772.
ACCESSION AR002596
VERSION AR002596.1 GI:3964150
```

AR002596 5087 bp DNA linear PAT 04-DEC-1998







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 1663.54 Seconds  
(without alignments)  
7973.437 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 819

Sequence: 1 tattattaaagcttcgcgg.....aggccagtcagctgtgctt 819

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	644.4	78.7	887	14	BQ948158
2	568.6	69.4	1157	14	BQ940483
3	544	66.4	1053	12	BG164929
4	543	66.3	853	12	BG437538
5	504.6	61.6	594	14	BM763333
6	491.2	60.0	580	14	BM848189

7	485.2	59.2	573	14	BM840863
8	476	58.1	476	10	BE632644
9	464.4	56.7	955	9	AL543945
10	447	54.6	522	14	BM841897
11	445	54.3	542	14	BM821005
12	441.4	53.9	529	14	BM847924
13	440.893	53.8	532	14	BM846370
14	407.6	49.8	569	14	BM845748
15	377.8	46.1	458	13	BM363136
16	351.2	42.9	420	14	BM764238
17	333.6	40.7	512	9	AI390475
18	329.8	40.3	482	14	BM846622
19	312.4	38.1	407	9	AI752561
20	263	32.1	291	10	BM846730
21	246.8	30.1	552	12	EG095271
22	246.693	30.1	488	9	AA015243
23	228.4	27.9	440	12	BG148676
24	228	27.8	913	13	BI912197
25	172.8	21.1	915	12	BF035982
26	155.8	19.0	560	9	AL800340
27	124.8	15.2	691	12	BF213570
28	111.8	13.7	633	13	BJ099801
29	102.6	12.5	655	13	BJ096761
30	66.4	8.1	180	9	AA204015
31	50.8	6.2	711	12	BG758365
32	50.2	6.1	925	17	CNS0091P
33	45.6	5.6	1309	17	AG077201
34	45	5.5	925	17	CNS0091P
35	43.4	5.3	853	14	BQ951739
36	43.4	5.3	936	17	CNS01SR2
37	43.4	5.3	1101	17	CNS01288
38	42.8	5.2	844	17	CNS0052P
39	42.8	5.2	932	17	CNS0072Q
40	42.4	5.2	505	9	AI984845
41	42.4	5.2	658	10	AW873554
42	42.4	5.2	685	13	BM005788
43	42.4	5.2	690	12	BF732971
44	42.4	5.2	787	12	BF971932
45	42.4	5.2	865	14	BQ672955

#### ALIGNMENTS

RESULT 1  
BQ948158  
LOCUS  
DEFINITION  
AGENCOURT\_8813192 NIH\_MGC\_101 Homo sapiens cDNA clone IMAGE:6428214  
5', mRNA sequence.  
ACCESSION  
BQ948158  
VERSION  
BQ948158.1  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 887)  
NIH-MGC <http://mgc.nhl.nih.gov/>.  
AUTHORS  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE  
Unpublished (1999)  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LLCM2614 row: h column: 07  
High quality sequence stop: 674.  
Location/Qualifiers  
1. .887

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6428214"  
/clone\_lib="NIH\_MGC\_101"  
/tissue\_type="epidermoid carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pOFB7; Site\_1: EcoRI; Site\_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 154 a 325 c 240 g 167 t 1 others

Query Match 78.7%; Score 644.4; DB 14; Length 887;  
Best Local Similarity 92.3%; Pred. No. 3.6e-142;  
Matches 700; Conservative 0; Mismatches 56; Indels 2; Gaps 2;

QY 2 ATTATTAAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCGAGCTCTGGGAGAG 61  
DB 5 ATTATTAAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCGAGCTCTGGGAGAG 64  
QY 62 GAGCG 121  
DB 65 GAGCG 123  
QY 122 GGGATGTTAGCTTGGCTATGACGGTGTGTGGCAGCTCCCTCGAGTGCAGCTCTTAAT 181  
DB 124 GGGATGTTAGCTTGGCTATGACGGTGTGTGGCAGCTCCCTCGAGTGCAGCTCTCAAT 183  
QY 182 CGCACAGGAGATCCAGCGCTGCGCCCTCCATCCAGAAAACTATGACCTACCCGCTAC 241  
DB 184 CGCACAGGAGATCCAGCGCTGCGCCCTCCATCCAGAAAACTATGACCTACCCGCTAC 243  
QY 242 CTGGAGCATCACTCGCAGCTTAGTGGGACCTACTGACCTACCTGGGCGCCCTTTC 301  
DB 244 CTGGAGCATCACTCGCAGCTTAGTGGGACCTACTGACCTACCTGGGCGCCCTTTC 303  
QY 302 AACGAGCTGACTTCAATCTCTCTCGAC-TGGGGGCGAGAACTCTGCCAGGGCCACGCT 360  
DB 304 AACGAGCATCACTCGCAGCTTAGTGGGACCTACTGACCTACCTGGGCGCCCTTTC 363  
QY 361 CAATTTGGAAGTGTGGGAGACCTCAATGACAGGCTCGGCTGACCCAGCACTATGAGGC 420  
DB 364 TGACTTGGAGTGTGGGAGACCTCAATGACAAACTCGGCTGACCCAGAACTACGAGGC 423  
QY 421 GTACAGTCACTCTCTGTGTTACTTGGCTGCGCTCAACCGTCAGGCTGCCAGCTGAAC 480  
DB 424 CTACAGCACCTCTCTGTGTTACTTGGCTGCGCTCAACCGTCAGGCTGCCAGCTGAAC 483  
QY 481 CCGACGTAGCTTGCGCCCACTTCTGTACAGCTCCAGGGCTGTGGGACGATTCGAGG 540  
DB 484 GCGCCGAGCTTGCGCCCACTTCTGTACAGCTCCAGGGCTGTGGGACGATTCGAGG 543  
QY 541 TGTGATGGAGCGCTTGCGCTACCACTGCGCCAGCTCTGCCAGGACTGAGCCAGCTG 600  
DB 544 CGTCATGGAGCTTGCGCTACCACTGCGCCAGCTCTGCCAGGACTGAGCCAGCTG 603  
QY 601 GGGCCCTGGCCCTGCCACAGTGAATCTCTCCAGAAATGATGATCTCTGGCTGCTGAA 660  
DB 604 GACTCTGGCCCTGCCACAGTGAATCTCTCCAGAAATGATGATCTCTGGCTGCTGAA 663  
QY 661 GGAGCTCAGACCTGGCTGTAGCGTTAGCGAAGCAAGCACTTCAACCGGCTTAGAAGAAAT 720  
DB 664 GGAGCTCAGACCTGGCTGTAGCGTTAGCGAAGCAAGCACTTCAACCGGCTTAGAAGAAAT 723  
QY 721 GCAGCTCCAGACCTTCAGTCACCTTCAGCTGAGG 758  
DB 724 GCAGCTCCAGACCTTCAGTCACCTTCAGCTGAGG 761

RESULT 2  
BQ940483  
LOCUS  
DEFINITION  
AGENCOURT 8864294 Lupski\_sciatic\_nerve Homo sapiens cDNA clone  
IMAGE:6197786 5', mRNA sequence.  
BQ940483  
BQ940483.1 GI:22355961  
EST.  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1157)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgraphs@email.nih.gov  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNL at:  
http://image.llnl.gov  
Plate: LLAM13608 row: g column: 03  
High quality sequence stop: 572.  
Location/Qualifiers  
1. 1157  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6197786"  
/sex="male"  
/tissue\_type="sciatic nerve"  
/dev\_stage="adult, 70 yr"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1:  
NotI; Site 2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCCAGCGTCCG-3' and  
5'-GACTAGTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >  
1 kb for average length 1.87 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine) and is available through Life  
Technologies."

BASE COUNT 221 a 374 c 346 g 216 t

Query Match 69.4%; Score 568.6; DB 14; Length 1157;  
Best Local Similarity 89.8%; Pred. No. 3.4e-124;  
Matches 622; Conservative 0; Mismatches 69; Indels 2; Gaps 1;

QY 109 AGGGAGCTCTGGGGATGTTAGCTTGCCTATGACGGTCTGTGGACCTCCCTGCAGT 168  
DB 58 AGGGAGCTCTGGGGATGTTAGCTTGCCTATGACGGTCTGTGGACCTCCCTGCAGT 117  
QY 169 GCAGCTCTTATCCACAGGAGATCCAGGCTGCGCCCTCCATCCAGAAACCTATGA 228  
DB 118 GCAGCTCTTATCCACAGGAGATCCAGGCTGCGCCCTCCATCCAGAAACCTATGA 177  
QY 229 CCTCACCCTGCTGGAGCATCACTCCGAGCTTGTAGTGGGACCTTACCTGAACTACCT 288  
DB 178 CCTCACCCTGCTGGAGCATCACTCCGAGCTTGTAGTGGGACCTTATCTGAACTACCT 237  
QY 289 GGGGCCCCCTTCAACAGGAGCTTCAATCTCTCTCGACTGGGGGAGAACTCTGCCC 348  
DB 238 GGGGCCCCCTTCAACAGGAGCTTCAATCTCTCTCGACTGGGGGAGAACTCTGCCC 297  
QY 349 CAGGGCCAGCTCAACTTGGAGTGTGGGAGAGCTCAATGACAGGCTGGGGCTGAGCCA 408

```

Db 298 CAGGCCACTGTTGACTTGGAGGTGTGGGAAAGCCTCAATGACAACTGCGGCTGACCCA 357
QY 409 GAATATAGAGCGGTACAGTCACTCCCTGTTACTTGGTGGCCTCAACCGTCAAGGCTGC 468
Db 358 GAATACAGAGCGGTACAGCAGCCTTCTGTGTTACTTGGTGGCCTCAACCGTCAAGGCTGC 417
QY 469 CACAGCTGAATCTCGAGTACGCTGGCCCACTTCTGTACAGAGCCTCCAGGCGCTGCTGGG 528
Db 418 CACTGTGAGTGTGGCCGCGAGCCTTGGCCCACTTCTGTACAGAGCCTCCAGGCGCTGCTGGG 477
QY 529 CAGCATTTGAGGTGTGATGGCGAGCTTGGCTACCCCTACCCAGCCTCCAGGCGCTGCTGGG 588
Db 478 CAGCATTTGCGGGCTCATGAGCAGCTTGGGCTACCCCACTGCGCCAGCGCTGCGCTGGGAC 537
QY 589 TGAGCCAGCTGGGCGCCCTGGCCCTGCCCCAAGTCACTTCTCCAGAGAGTGGATGACTT 648
Db 538 TGAACCCACTTGGACTCTGGCCCTGCCCCAAGTCACTTCTCCAGAGAGTGGATGACTT 597
QY 649 CTGGCTGTCTGAAGAGCTGAGAGCTGGCTATGGCGTTAGCGTTTCAAGAGAGTGGATGACTT 708
Db 598 CTGGCTGTCTGAAGAGCTGAGAGCTGGCTATGGCGTTAGCGTTTCAAGAGAGTGGATGACTT 657
QY 709 TAAGAGAGATGAGAGCTGAGAGCTGGCTATGGCGTTAGCGTTTCAAGAGAGTGGATGACTT 766
Db 658 CAAGAGAGATGAGAGCTGAGAGCTGGCTATGGCGTTAGCGTTTCAAGAGAGTGGATGACTT 717
QY 767 TTCTGACTCTGACCTTACCCCTTACCCCTTACCCCTTACCCCTTACCCCTTACCCCTT 799
Db 718 TTCTGACTCTGACCTTACCCCTTACCCCTTACCCCTTACCCCTTACCCCTTACCCCTT 750

RESULT 3
BG164929
LOCUS
DEFINITION 60234555F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4453813 5',
mRNA sequence.
ACCESSION BG164929
VERSION BG164929.1 GI:12671563
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0244 row: m column: 14
High quality sequence start: 3
High quality sequence stop: 675.
Location/Qualifiers
1..1053
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4453813"
/clone_lib="NIH_MGC_89"
/tissue_type="hypermorphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: Kidney; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 210 a 351 c 292 g 200 t

```

## ORIGIN

```

Query Match 66.4%; Score 544; DB 12; Length 1053;
Best Local Similarity 88.2%; Pred. No. 2.2e-118;
Matches 660; Conservative 0; Mismatches 80; Indels 8; Gaps 6;

QY 24 CGCGGCTGCGCTCCCACTCGCCAGCCTCTGGGAGAGGAGCCGCGCCGCGCCGCGG 83
Db 2 CGCGGCTGCGCTCCCACTCGCCAGCCTCTGGGAGAGGAGCCGCGCCGCGCCGCGG 61
QY 84 CCCCCAGCCCCATGGAGCCTCCGAGCAGGGGACTCTGGGGGAGTGTAGCTTGCCTATGCA 143
Db 62 CCCCCAGCCCCATGGAGCCTCCGAGCAGGGGACTCTGGGGGAGTGTAGCTTGCCTATGCA 121
QY 144 CGGTGCTGTGGCACTCCCTCCCTGCACTGCACTCTTAATCGCAGAGAGATCCAGGCGCTG 203
Db 122 CGGTGCTGTGGCACTCCCTCCCTGCACTGCACTCTTAATCGCAGAGAGATCCAGGCGCTG 181
QY 204 GCCCCTCCATCCAGAAACCTATGACCTCACTCCCTGCACTGCACTGCACTGCACTGCACT 263
Db 182 GCCCCTCCATCCAGAAACCTATGACCTCACTCCCTGCACTGCACTGCACTGCACTGCACT 241
QY 264 TAGCTGGGACCTACCTGAACTACCTGGGGCCCCC---TTTCAACGAGCCT-GACTTCAAT 319
Db 242 TGCTGGGACCTATCTGAATAGCTTGGGGCCCCCCTTTCACTGAGCCTTAGACTTCAAC 301
QY 320 CCTCTCGACTGGGGGCGAGAACTCTG-CCAGGGCCACGGTCAACTTGGAACTGTGGCG 378
Db 302 CCTCCCGCTGGGGGCGAGAACTCTGTCCTCCAGGGCCACTGTTTGAATTTGGAGTGTGGCG 361
QY 379 AAGCCTCAATGACAGGCTCGGCTGACCCAGAACTATGAGGGGTACAGTCACTCTCTGTG 438
Db 362 AAGCCTCAATGACAACTCGGCTGACCCAGAACTACAGAGCCTACAGCCACCTCTCTGTG 421
QY 439 TTACTTGGTGCTTCAACCGTCACTGAGCTGCCAGCTGAACTCCGAGTACCTGCGCCCA 498
Db 422 TTACTTGGTGCTTCAACCGTCACTGAGCTGCCAGCTGAGCTGCGCCGAGCCTGCGCCCA 481
QY 499 CTTCTGATACAGCCTCCAGGGCTCTGCGGAGAGATTTGAGGTGTCATGGCACTGTGG 558
Db 482 CTTCTGATACAGCCTCCAGGGCTCTGCGGAGAGATTTGCGGCGCTCATGGCAGCTCTGGG 541
QY 559 CTACCACTG-CCCCAGGCTCTGCGGAGGAGTGTG-AGCCAGCCTGCGGCGCTGCGCCCTGCG 616
Db 542 CTACCACTGCGCCCGAGCGCTGCTGCGGAGTGTGAACTGAACTGCTTGGACTCTGCGCCCTGCG 601
QY 617 CACAGTGAATCTTCCAGAGAGATGATGATCTTGGCTGCTGAAAGAGAGTGCAGACCTGG 676
Db 602 CACAGTGAATCTTCCAGAGAGATGATGATCTTGGCTGCTGAAAGAGAGTGCAGACCTGG 661
QY 677 CTATGGCGTTCCAGCAGGAGCTTCAACCGGCTTAAGAGAGAGATGCAGCTTCCAGAGCT 736
Db 662 CTGTGGCGCTGGCG-AGGACTTCCACCGTCCAGAGAGATTTGAGCCTTCCAGAGCTT 720
QY 737 TCATGACCTTGCACCTTGGAGGCGCATG 764
Db 721 GCAGTCACTGGACCTGGGGGGCCATG 748

RESULT 4
BG437538
LOCUS
DEFINITION 60248927F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4621530 5',
mRNA sequence.
ACCESSION BG437538
VERSION BG437538.1 GI:13344044
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/

```

Db	539	CCACTTGGACTCTCTGGCCCTTGCCACAGTGA	CTTCTCCAGAAAGATGGACGACTTCTGGC	598
QY	654	TGCTGAAGAGCTGCAGACTGCCTATAGCGTT	TACGCAAGGACTTCAA-CCGCGCTTAAG	712
Db	599	TGCTGAAGAGCTGAAGACTGCTGCTGTGGCG	CTGGCCACGAGACTTCAAGCCGCTCAAG	658
QY	713	AGAAAGATGCAGCTCTCAG-CAGCTTTCAGT	CACCC---TGCACTTGGAGGCAATGGTTT	768
Db	659	AGAAAGATGCAGCTCTCAGCCACCTGCAGT	CAACCCATGCACCTTGGAGGCTCATGGGTT	718
QY	769	CTGACCTCTGACCTTAACCCCCACACCTCC	799	
Db	719	CTGACTTCTGACCTTCTCTCTCTTCGCTCC	749	

RESULT 5

BM763333

LOCUS

DEFINITION

BM763333

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

598

712

658

768

718

799

749

101 a

226 c

164 g

103 t

61.6%

Score 504.6;

DB 14;

Length 594;

91.6%;

Pred. No. 3.9e-109;





Best Local Similarity 100.0%; Pred. No. 2.1e-102; Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 GGGCCCTCCATCCAGAAACCTATGACCTCACCCTGCTACCTGGAGCATCAACTCCGAGC 262  
 Db 1 GGGCCCTCCATCCAGAAACCTATGACCTCACCCTGCTACCTGGAGCATCAACTCCGAGC 60

QY 263 TTAGTGGGACCTACTGAACTACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCCT 322  
 Db 61 TTAGTGGGACCTACTGAACTACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCCT 120

QY 323 CCTGACTGGGGGAGAACTCTGCCAGGGCCAGCTCACTGGAGTGGGGAAGC 382  
 Db 121 CCTGACTGGGGGAGAACTCTGCCAGGGCCAGCTCACTGGAGTGGGGAAGC 180

QY 383 CTCATGACAGCTCGGGCTGACCCAGAACTATGAGGGCTACAGTCACCTCTGTGTAC 442  
 Db 181 CTCATGACAGCTCGGGCTGACCCAGAACTATGAGGGCTACAGTCACCTCTGTGTAC 240

QY 443 TTGGTGGCCCTCAACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 502  
 Db 241 TTGGTGGCCCTCAACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 300

QY 503 TGTACAGCTCCAGGGCTGTGGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 562  
 Db 301 TGTACAGCTCCAGGGCTGTGGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 360

QY 563 CCATGCCCCAGCTCTGCCAGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 622  
 Db 361 CCATGCCCCAGCTCTGCCAGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 420

QY 623 GACTTCTCCAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 678  
 Db 421 GACTTCTCCAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 476

RESULT 9  
 AL543945  
 LOCUS  
 DEFINITION AL543945 LTI\_NFL006\_P12 Homo sapiens cDNA clone CS0D1004YML5 5 prime, mRNA sequence.  
 VERSION AL543945  
 KEYWORDS AL543945.1 GI:12876424  
 SOURCE EST.  
 ORGANISM human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
 source  
 1..955  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1004YML5"  
 /clone\_lib="LTI\_NFL006\_P12"  
 /tissue\_type="placenta"  
 /notes="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 168 a 323 c 266 g 198 t

Query Match 56.7%; Score 464.4; DB 9; Length 955;  
 Best Local Similarity 90.5%; Pred. No. 1.5e-99;  
 Matches 507; Conservative 0; Mismatches 51; Indels 2; Gaps 1;

QY 224 TATGACCTCACCCGCTACCTGGAGCATCAACTCCGAGCTTACCTGGGACCTTACCTGAAC 283  
 Db 1 TATGACCTCACCCGCTACCTGGAGCATCAACTCCGAGCTTACCTGGGACCTTATCTGAAC 60

QY 284 TACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCTCTCTCGACTGGGGGAGAACT 343  
 Db 61 TACCTGGGGCCCCCTTTCAACGAGCAGACTTCAACCTCTCCCGCTGGGGGAGAGACT 120

QY 344 CTGCCAGGGCCAGGTCACCTTGAAGTGGGAGAGCTCAATGACAGGCTGGGGCTG 403  
 Db 121 CTGCCAGGGCCAGTGTGAGCTTGGAGGTGGGCGAA--CTCAATGACAAACTGGGGCTG 178

QY 404 ACCAGAACTATGAGGGCTACAGTCACCTCTCTGTACTTGTGCGTGGCTCAACCGTCAAG 463  
 Db 179 ACCAGAACTACGAGGCTACAGCCACCTTCTGTGTACTTGGTGGCTCAACCGTCAAG 238

QY 464 GCTGCCAGCTGAACTCCGAGCTAGCCTGGCCCACTTCTGTACAGGCTCCAGGGCTG 523  
 Db 239 GCTGCCAGCTGAGCTGGCGGAGCCTGGCCCACTTCTGTACAGGCTCCAGGGCTG 298

QY 524 CTGGGAGCATTTGAGGCTGATGCGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 583  
 Db 299 CTGGGAGCATTTGCGGGCTCATGGGAGCTCTGGGCTTACCCAGCTGGCGGCTGGCT 358

QY 584 GGGACTGAGCCAGCTGGGGCCCCCTGGCCCTGGCCAGCTGACTTCTCTCAGAAAGATGAT 643  
 Db 359 GGGACTGAAACCCACTTGGAGCTCTGGCCCTGGCCAGCTGACTTCTCTCAGAAAGATGAT 418

QY 644 GACTTCTGGCTGCTGAAGGAGCTGGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 703  
 Db 419 GACTTCTGGCTGCTGAAGGAGCTGGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 478

QY 704 CGGCTTAAGAAGAGAGTCCAGCTCCAGCAGCTTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 763  
 Db 479 CGGCTTAAGAAGAGAGTCCAGCTCCAGCAGCTTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 538

QY 764 GGTTCCTGACCTCTGACCT 783  
 Db 539 GGTTCCTGACCTCTGACCT 558

RESULT 10  
 BM841897  
 LOCUS  
 DEFINITION BM841897 522 bp mRNA linear EST 06-MAR-2002  
 K-EST0119216 SL3KMS5 Homo sapiens cDNA clone SL3KMS5-34-A05 5', mRNA sequence.  
 ACCESSION BM841897  
 VERSION BM841897.1 GI:19198306  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 522)  
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R., Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and Kim, Y.S.  
 21C Frontier Korean EST Project 2001  
 Unpublished (2002)  
 Contact: Kim YS  
 Genome Research Center  
 Korea Research Institute of Bioscience & Biotechnology  
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
 Tel: +82-42-860-4470  
 Fax: +82-42-860-4409  
 Email: yongsung@mail.kribb.re.kr

---



```

Db 237 ACTCCGACCTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAACGAGCAGA 296
QY 313 CTTCAATCTCTCGACTGGGGGCGAAGAACTCTGCCAGGCGCACGGTCAACTTTGGAAGT 372
Db 297 CTTCAACCTCTCCCGCTGGGGGCGAGACTCTGCCAGNGCCACTGTGTGACTTGGAGGT 356
QY 373 GTGGGGAAGCCTCAATGACAGGCTGGGTGACCCAGAACTATGAGGCGTACAGTCACT 432
Db 357 GTGGGGAAGCCTCAATGACAAACTTGGCGGTGACCCAGAACTACGAGGCGTACAGCCACT 416
QY 433 CTTGTGTACTTGGCTGGCTCAACCGTCAAGCTGCGACAGCTGCACTCGAGCTAGCCT 492
Db 417 TCTGTGTACTTGGCTGGCTCAACCGTCAAGCTGCGACAGCTGCTGCTGCGCGAGCCT 476
QY 493 GCGCCACTTCTGTACAGCCTCCAGGGCTGCTGGGCGACATTCAGAGTGTCTATGGCGAC 552
Db 477 GCGCCACTTCTGCACAGCCTCCAGGGCTGCTGGGCGACATTCGGGCGTCTATGGCAGC 536
QY 553 GCTTGG 558
Db 537 TCTGGG 542

RESULT 12
BM847924 529 bp mRNA linear EST 06-MAR-2002
LOCUS K-EST0127535 S13KMS5 Homo sapiens cDNA clone S13KMS5-57-A03 5',
DEFINITION mRNA sequence.
ACCESSION BM847924
VERSION BM847924.1 GI:19204323
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 529)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 57 row: A column: 03
High quality sequence stop: 529.
Location/Qualifiers
1. 529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S13KMS5-57-A03"
/clone_lib="S13KMS5"
/tissue_type="myeloma"
/cell_line="RMS-5"
/lab_host="Top10P"
/notes="Vector: pCNS; Site_1: EcoRI; Site_2: NotI; The poly
(A)+ RNA was dephosphorylated with bacterial alkaline
phosphatase (BAP) and then decapped with tabacco acid
pyrophosphatase (TAP). The decapped intact mRNA was
ligated with DNA-RNA linker including EcoR I site by
treatment of T4 RNA ligase and the first strand cDNA was
synthesized from oligo dT-selected mRNA by priming with
dT-tailed vector. The dT-tailed vector was adjusted to
have about 60nt. The cDNA vector was circularized with E.
coli DNA ligase after digestion of EcoRI which site is
also included in vector. An RNA strand converted to a DNA
strand by Okayama-Berg method. The obtained cDNA vectors
were used for transformation of competent cells E. coli

```

```

Top10P' by electroporation method. The cDNA libraries
constructed by this method are full-length enriched cDNA
library."
BASE COUNT 92 a 199 c 147 g 91 t
ORIGIN
Query Match 53.9%; Score 441.4; DB 14; Length 529;
Best Local Similarity 91.9%; Pred. No. 3.4e-94;
Matches 488; Conservative 0; Mismatches 41; Indels 2; Gaps 2;
QY 35 CTCCTCCCTCGCAGCTCTGGGAGAGAGCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 94
Db 1 CTCCTCCCTCGCAGCTCTGGGAGAGAGCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 59
QY 95 ATGGACCTCCGAGAGGAGGACTCGTGGGGATGTTAGCTTGGCTATGACGGTGTGTGG 154
Db 60 ATGGACCTCCGAGCA-GGAGCTCGTGGGGATGTTAGCGTGTGTGACGGTGTGTGG 118
QY 155 CACCTCCCTCGCAGTGCAGCTCTTAATGCGACAGAGAGATCCAGGGCCTGGCCCTCCATC 214
Db 119 CACCTCCCTCGCAGTGCAGCTCTCAATCGCACAGGGACCCAGGGCTGGCCCTCCATC 178
QY 215 CAGAAACCTATGACCTCACCGCTACCTGGAGCATCAACTCCGAGCTTAGCTGGGACC 274
Db 179 CAGAAACCTATGACCTCACCGCTACCTGGAGCACCAACTCCGAGCTTAGCTGGGACC 238
QY 275 TACCTGAACCTACCTGGGCGCCCTTTCAAGAGCCTGACTTCAATCCTCTCGACTGGGG 334
Db 239 TATCTGAACCTACCTGGGCGCCCTTTCAAGAGCCTGACTTCAATCCTCTCGACTGGGG 298
QY 335 GCAGAACTCTGCCAGGGCCACCGTCAACTTGGAGTGTGGCGAAGCTCAATGACAGG 394
Db 299 GCAGAGACTCTGCCAGTGCACCTGTTCGCTTGGAGTGTGGCGAAGCTCAATGACAAA 358
QY 395 CTGGGCTGACCCAGAACTATGAGCGTACGTACCTCCCTGTGTGTACTTGGCTGGCTTC 454
Db 359 CTGGGCTGACCCAGAACTATGAGCGCTACAGCCACCTTCTGTGTACTTGGCTGGCTTC 418
QY 455 AACCTGAGGCTGCCACAGCTGAACTCCGAGTACCTGGCGCCACTTCTGTACCGAGCTTC 514
Db 419 AACCTGAGGCTGCCACAGCTGAACTCCGAGTACCTGGCGCCACTTCTGTACCGAGCTTC 478
QY 515 CAGGCTGCTGGCGAGCATTCGAGGTGTATGGCGAGCTTGGCTAGCCCA 565
Db 479 CAGGCTGCTGGCGAGCATTCGAGGTGTATGGCGAGCTTGGCGGTATCCCA 529

RESULT 13
BM846370
LOCUS K-EST0125376 S13KMS5s1 Homo sapiens cDNA clone S13KMS5s1-21-A09 5',
DEFINITION mRNA sequence.
ACCESSION BM846370
VERSION BM846370.1 GI:19202769
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 532)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 21 row: A column: 09

```



```

165 GAGCGCACCCGCGCGCGCG-CCCCAGCCCCATGGACCTCCGAGCAGGGGACTCGTGG 223
122 GGGATGTTAGCTTGGCTATGACAGGTGTGTGGACCTCCCTGACAGTCCAGCTCTTAAT 181
224 GGGATGTTAGCTTGGCTATGACAGGTGTGTGGACCTCCCTGACAGTCCAGCTCTTAAT 283
182 CGCAGAGAGATCCAGGCGCTTGGCGCTTCCATCCAGAAAACTATGACCTCACCGCTAC 241
284 CGCAGAGAGATCCAGGCGCTTGGCGCTTCCATCCAGAAAACTATGACCTCACCGCTAC 343
242 CTGAGGATCAACCTCCGAGCTTGGGACCTTACCTGAGTACCTGAGGCGCGCTTTC 301
344 CTGAGGATCAACCTCCGAGCTTGGGACCTTACCTGAGTACCTGAGGCGCGCTTTC 403
302 AACGAGCTGATCTCAATCTCTGACTGGGGGAGAAAACTTGGCGCGCGCTTTC 361
404 AACGAGCTGATCTCAATCTCTGACTGGGGGAGAAAACTTGGCGCGCGCTTTC 463
362 AACTTGAAGTGTGGCGAGCTCAATGACAGGTGTGGGCTGACCCAGAACTATGAGCG 421
464 GACTTGAAGTGTGGCGAGCTCAATGACAACTTGGGCTGACCCAGAACTACGAGGCC 523
422 TACAGTCACTCTGTGTGTTACTTGGGCTTCAACCTCAGGCTG 467
524 TACAGCACCCTTGTGTGTTACTTGGGCTTCAACCTCAGGCTG 569

```

```

RESULT 15
BM363136
LOCUS BM363136 458 bp mRNA linear EST 10-JAN-2002
DEFINITION BS320054A10B08 Subtracted Lewin Cattle Spleen cDNA clone
BM363136
ACCESSION BM363136.1 GI:18106505
VERSION EST.
KEYWORDS cow.
SOURCE Bos taurus
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1. (bases 1 to 458)
Lewin,H.A., Soares,M.B., Pardinas,J., Liu,L. and Larson,J.H.
Subtracted Lewin Cattle Spleen ESTs
Unpublished (2002)
Contact: Lewin, H. A.
W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Funding for Cattle EST sequencing was provided by the USDA National
Research Initiative, Project No. 98-35205-6644, and a grant from
the Japanese Ministry of Agriculture Fisheries and Forestry to
H.A.Lewin and J.E.Womack. Base-calling/Quality scores: PHRED form
Washington University Genome Center. Vector-trimming: Cross Match
from Washington University Genome Center PHRAP suite. This sequence
is vector free and at least 200bp in length.
Insert length: 458 Std Error: 0.00
Plate: BS320054A10 row: E column: 08
Seq primer: CGCAGGCTCGAAATTAACCC
High quality sequence stop: 458.
Location/Qualifiers
1. .458
/organism="Bos taurus"
/strain="Angus"
/db_xref="taxon:9913"
/clone="BS320054A10E08"
/clone_lib="Subtracted Lewin Cattle Spleen"
/sex="Female"
/dev_stage="Adult"

```

## FEATURES

source

/note="Organ: Spleen; Vector: pBluescript SK(+/-); Site 1: EcoRI; Site 2: XhoI; This BS cDNA library was generated by subtraction of the original non-normalized clones from a library with 16,800 previously sequenced clones from a bovine Placenta cDNA library. The original library was constructed as described by Band et al (2000), Genome Research 10(9): 1359-1368."

BASE COUNT 82 a 182 c 125 g 69 t

Query Match 46.1%; Score 377.8; DB 13; Length 458;  
 Best Local Similarity 89.6%; Pred. No. 3.6e-79;  
 Matches 406; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 187 AGGAGATCCAGGCGCTTGGCGCTTCCATCCAGAAAACTATGACCTCACCGCTACCTCGA 246  
 DB 6 AGGAGATCCAGGCGCTTGGCGCTTCCATCCAGAAAACTATGACCTCACCGCTACCTCGA 65  
 QY 247 GCATCAATCCGACGCTTAGCTGGGACCTTACCTGAACTTACCTGGGGCCCCCTTTCAACGA 306  
 DB 66 GCACGAGCTCGGACGCTTGGCTGGGACCTTCTGAACTTACCTGGGGCCCCCTTTCAACGA 125  
 QY 307 GCCTGACTTCAATCTCTCTGACTGGGGGAGAACTTGTGCCAGGGGCCAGGCTCAACTT 366  
 DB 126 GCCTGACTTCAACCCCCCTCTGGCTGGGGGTGGAGACTTGTGCCAAGGCCACTGTCAACT 185  
 QY 367 GGAAGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGGCTACAG 426  
 DB 186 GGAGTGTGGCGAAGCCTCAACGAACTTGGCTGACCCAGAACTTACGAGGGCTACAG 245  
 QY 427 TCACCTCTCTGTGTTACTTGGCTGGGCTTCAACGCTCAGGCTGCCACAGCTGAACTCCGACG 486  
 DB 246 CCACCTCTCTGTGCTACCTTGGCGGCTTCAACGCGCAGGCGCCACAGCGGAGCTGGCGG 305  
 QY 487 TAGCTGGCGCACTTCTGTACAGGCTTCCAGGGCTGCTGGGGAGCAATGCAAGTGTCTAT 546  
 DB 306 CAGCTTGGCGCACTTCTGTACAGGCTTCCAGGGCTGCTGGGGAGCAATGCAAGTGTCTAT 365  
 QY 547 GGCAGCTTGGCTTACCCACTTGGCGGCTTCCAGGGCTTCCAGGGGACTGAGCCAGCTGGGGCCC 606  
 DB 366 GGCAGCTTGGGCTTACCGCTTGGCGGCTTCCAGGGCTTCCAGGGGACTGAGCCAGCTGGGGCCC 425  
 QY 607 TGGCCCTTGGCCACAGTGAATCTTCTTCCAGAAAGAT 639  
 DB 426 TGGCCCTTGGCCACAGTGAATCTTCTTCCAGAAAGAT 458

Search completed: February 1, 2003, 12:59:17  
 Job time : 1669.54 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 208.568 Seconds  
(without alignments)  
8843.084 Million cell updates/sec

Title: US-09-931-704-4  
Perfect score: 819  
Sequence: 1 tattattaagcttcgcgg.....aggccagtcagctgtgctt 819

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002:\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	819	100.0	819	19 AAV47512	Mouse neurotrophic cDNA encoding murti
2	819	100.0	819	19 AAV22654	Murine NNT-1 cDNA.
3	819	100.0	819	21 AAA39483	Mouse cDNA encoding
4	819	100.0	819	22 AAA74484	Nucleotide sequenc
5	674	82.3	881	22 AAA74484	Human interleukin-
6	674	82.3	1790	21 AAA88546	Human neurotrophic cDNA encoding huma
7	669.4	81.7	797	19 AAV47510	Human NNT-1 cDNA.
8	669.4	81.7	797	19 AAV22652	
9	669.4	81.7	797	21 AAA39481	

10	669.4	81.7	797	24 ABK11647	Human cDNA encodin
11	660	80.6	768	22 AAH99772	Human protein enco
12	646.4	78.9	648	21 AAA88547	Mouse interleukin-
13	627	76.6	1710	20 AAX16161	Human cardiotroph
14	600.2	73.3	729	22 AAD04201	Human cardiotroph
15	583.2	71.2	1008	22 AAX51548	Human polynucleoti
16	577	70.5	968	22 ABA09140	Human polynucleoti
17	577	70.5	968	22 ABA09140	Human polynucleoti
18	429.6	52.5	5087	21 AAA39482	Human NNT-1 DNA.
19	429.6	52.5	5087	21 AAA39482	Human NNT-1 DNA.
20	429.6	52.5	5088	19 AAV47511	Human novel neurot
21	429.6	52.5	5088	19 AAV47511	Human novel neurot
22	419.8	51.3	495	22 ABA22653	Human neurotrophic
23	419.8	51.3	495	22 ABA22653	Human neurotrophic
24	419.8	51.3	495	22 ABA22653	Human neurotrophic
25	419.8	51.3	495	22 ABA22653	Human neurotrophic
26	419.8	51.3	495	22 ABA22653	Human neurotrophic
27	419.8	51.3	495	22 ABA22653	Human neurotrophic
28	419.8	51.3	495	22 ABA22653	Human neurotrophic
29	299.2	36.5	492	22 ABA22653	Human neurotrophic
30	299.2	36.5	492	22 ABA22653	Human neurotrophic
31	299.2	36.5	492	22 ABA22653	Human neurotrophic
32	299.2	36.5	492	22 ABA22653	Human neurotrophic
33	299.2	36.5	492	22 ABA22653	Human neurotrophic
34	299.2	36.5	492	22 ABA22653	Human neurotrophic
35	299.2	36.5	492	22 ABA22653	Human neurotrophic
36	236	28.8	283	22 ABA22653	Human neurotrophic
37	65	7.9	65	24 ABA22653	Human neurotrophic
38	64	7.8	1260	24 ABA22653	Human neurotrophic
39	64	7.8	1260	24 ABA22653	Human neurotrophic
40	52	6.0	396	20 AAX16162	Oligonucleotide fo
41	49.2	6.3	627	20 AAX60796	Human cardiotroph
42	47.4	5.8	627	20 AAX60796	Human cardiotroph
43	43.2	5.3	6803	18 AAT85473	Human DNAX interle
44	43.2	5.3	6803	18 AAT85473	Human DNAX interle
45	42.4	5.2	1578	24 ABA22653	Genomic hNET seque

## ALIGNMENTS

RESULT 1  
AAV47512  
ID AAV47512 standard; cDNA; 819 BP.

AC AAV47512;  
DT 09-NOV-1998 (first entry)  
DE Mouse neurotrophic factor NNT-1 cDNA.

NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;  
Alzheimer's disease; Parkinson's disease; Huntington's disease;  
amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
peripheral neuropathy; dystrophy; neural retina degeneration;  
common variable immunodeficiency; CVID; selective IGA deficiency;  
hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;  
therapy; ss.

OS Mus sp.

Key Location/Qualifiers  
CDS 95..772  
FT sig\_peptide /\*tag= a  
FT mat\_peptide /\*tag= b  
FT /\*tag= c

WO9833922-A1.

06-AUG-1998.

XX

PF	XX	02-FEB-1998;	98WO-US02363.	DB	421	GTACAGTCACTCTCTGTTACTTGGCTTCAACCGTCAGGCTGCCACAGCTGAAC	480
XX	XX	30-JAN-1998;	98US-0016534.	QY	481	CCGACGTAGCTGGCCCACTTCTGTACAGCTCCAGGCTCTGGGCGAGCATTTGCAGG	540
PR	XX	03-FEB-1997;	97US-0792019.	DB	481	CCGACGTAGCTGGCCCACTTCTGTACAGCTCCAGGCTCTGGGCGAGCATTTGCAGG	540
XX	XX	(AMGE-) AMGEN INC.		QY	541	TGTCATGGCGACGCTTGGCTACCCACCTGCCCCAGGCTCTGCCAGGACTTGAGCCAGCCTG	600
XX	PI	Chang M, Elliot GS, Sarmiento U, Senaldi G;		DB	541	TGTCATGGCGACGCTTGGCTACCCACCTGCCCCAGGCTCTGCCAGGACTTGAGCCAGCCTG	600
XX	XX	WPI; 1998-437475/37.		QY	601	GGCCCCCTGGCCCTCCACAGTACTTCTCCAGAGATGATGATCTTCTGCTGCTGAA	660
XX	XX	P-PSDB; AAV29716.		DB	601	GGCCCCCTGGCCCTCCACAGTACTTCTCCAGAGATGATGATCTTCTGCTGCTGAA	660
PT	XX	Newly isolated nucleic acid encoding human or murine neurotrophic		QY	661	GGAGCTGCAGACCTGGCTATGGCTTACGCCAAGGACTTCAACCGGCTTTAAGAAGAAGAT	720
PT	XX	factor NNT-1 - useful for treatment of neurological and		DB	661	GGAGCTGCAGACCTGGCTATGGCTTACGCCAAGGACTTCAACCGGCTTTAAGAAGAAGAT	720
PT	XX	immunological diseases or inflammation, also as vaccine adjuvant		QY	721	GCAGCTCCAGCAGCTTCAGTCACTCCCTGACCTGAGGACATGTTCTGACCTCTGAC	780
XX	XX	Disclosure; Fig 4; 120pp; English.		DB	721	GCAGCTCCAGCAGCTTCAGTCACTCCCTGACCTGAGGACATGTTCTGACCTCTGAC	780
CC	CC	This newly isolated mouse cDNA sequence codes for a novel		QY	781	CCTTAACCCCACTCCAGCTCCAGGCCCCAGCTGCTGCTT	819
CC	CC	neurotrophic factor, designated NNT-1 (see AAV29716), that is a		DB	781	CCTTAACCCCACTCCAGCTCCAGGCCCCAGCTGCTGCTT	819
CC	CC	growth factor for neurons and for B or T cells. Vectors and		QY	819	CDNA encoding murine neurotrophic factor NNT-1.	
CC	CC	host cells are provided for use in the production of murine and		DB	819	Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;	
CC	CC	human NNT-1 polypeptides. These are used to treat: (i) neurological		QY	819	treatment; neurological disease; degeneration; Parkinson's disease;	
CC	CC	or immunological diseases, specifically Alzheimer's, Parkinson's or		DB	819	amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke; ss.	
CC	CC	Huntington's diseases, amyotrophic lateral sclerosis,		QY	819	Mus sp.	
CC	CC	Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and		DB	819	Key	
CC	CC	degeneration of the neural retina, or conditions characterised by T		QY	819	CDS	
CC	CC	cell defects, e.g. common variable immunodeficiency (CVID),		DB	819	/*tag= a	
CC	CC	selective IgA deficiency, hypogammaglobulinaemia and X-linked		QY	819	sig_peptide	
CC	CC	agammaglobulinemia (claimed), but many others disclosed; and (ii)		DB	819	/*tag= b	
CC	CC	inflammation. NNT-1 is also able to boost immunoreactivity and		QY	819	mat_peptide	
CC	CC	antibody production following vaccination, and, since it inhibits		DB	819	/*tag= c	
CC	CC	tumour necrosis factor production, it may also be useful for		QY	819	US5741772-A.	
CC	CC	treating sepsis. NNT-1 nucleic acid fragments are also used as		DB	819	21-APR-1998.	
CC	CC	hybridisation probes in diagnostic assays. In addition, cells that		QY	819	03-FEB-1997;	
CC	CC	have been engineered to express NNT-1 can be implanted, or nucleic		DB	819	03-FEB-1997;	
CC	CC	acids are delivered in gene therapy vectors.		QY	819	(AMGE-) AMGEN INC.	
XX	XX	Sequence 819 BP; 156 A; 288 C; 218 G; 157 T; 0 other;		DB	819	Chang M;	
XX	XX	Query Match 100.0%; Score 819; DB 19; Length 819;		QY	819	WPI; 1998-260526/23.	
XX	XX	Best Local Similarity 100.0%; Pred. No. 6e-196;		DB	819	P-PSDB; AAV56142.	
XX	XX	Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		QY	819	Neurotrophic factor NNT-1 polypeptide and related nucleic acids -	
QY	1	TATTATTAAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCGAGCTCTGGGAGA	60	PT	819	useful for stimulating growth of motor and sympathetic neurons	
DB	1	TATTATTAAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCGAGCTCTGGGAGA	60	PT	819	Disclosure; Fig 4; 41pp; English.	
QY	61	GGAGCG	120	XX	819	The present sequence encodes a murine neurotrophic factor, designated	
DB	61	GGAGCG	120	XX	819		
QY	121	GGGATGTTAGCTTGCCTATGACCGGCTGTGGGACCTCCCTGCGAGTCCAGCTCTTAA	180	XX	819		
DB	121	GGGATGTTAGCTTGCCTATGACCGGCTGTGGGACCTCCCTGCGAGTCCAGCTCTTAA	180	XX	819		
QY	181	TCGCACAGGATCCAGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	240	XX	819		
DB	181	TCGCACAGGATCCAGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	240	XX	819		
QY	241	CCTGGAGCATCAACTCCGAGCTTACCTGGGACCTACCTGAACCTACCTGGGCGCCCTTT	300	XX	819		
DB	241	CCTGGAGCATCAACTCCGAGCTTACCTGGGACCTACCTGAACCTACCTGGGCGCCCTTT	300	XX	819		
QY	301	CAACGAGCTGACTTCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	360	XX	819		
DB	301	CAACGAGCTGACTTCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	360	XX	819		
QY	361	CAACTTGGAAAGTGTGGGAGGCTCAATGACAGGCTCGGCTGACCCAGAACTATGAGGC	420	XX	819		
DB	361	CAACTTGGAAAGTGTGGGAGGCTCAATGACAGGCTCGGCTGACCCAGAACTATGAGGC	420	XX	819		
QY	421	GTACAGTCACTCTCTGTTACTTGGCTTCAACCGTCAGGCTGCCACAGCTGAAC	480	XX	819		

[illegible][illegible]

RESULT 3  
AAA39483  
ID AAA3

KN	grate
XX	
OS	Mus sp.

QY	1	TATTTATTAAGCTT	CGCCGGAGCGCGGCT	CGCCCTCCCACT	CCGCGCAGCTCT	GGGAGA	60
Db	1	TATTTATTAAGCTT	CGCCGGAGCGCGGCT	CGCCCTCCCACT	CCGCGCAGCTCT	GGGAGA	60
QY	61	GGAGCGCGCGCC	CGGCGCCGAGCCCC	ATGGACCT	CCGAGCAGGGGAC	TCGTCG	120
Db	61	GGAGCGCGCGCC	CGGCGCCGAGCCCC	ATGGACCT	CCGAGCAGGGGAC	TCGTCG	120
QY	121	GGGGATTTAGCTT	GTGCATGACAGGCT	GTGTGGACCT	CCCTGCAGTGC	CAGCTCTTAA	180
Db	121	GGGGATTTAGCTT	GTGCATGACAGGCT	GTGTGGACCT	CCCTGCAGTGC	CAGCTCTTAA	180
QY	181	TCGCACAGGAGAT	CCAGGGCCCT	TGGCCCTCCAT	TCCAGAAA	CCATTGACCT	240



Db 181 TCGACAGGAGATCCAGGCGCTGCGCCCTCCATCCAGAAAACCTATGACCTCACCCGCTA 240  
 QY 241 CTTGGAGCATCAACTCCGCGAGCTTAGCTGGGACCTACTGAACCTACTGAGGCGCCCTTT 300  
 Db 241 CTTGGAGCATCAACTCCGCGAGCTTAGCTGGGACCTACTGAACCTACTGAGGCGCCCTTT 300  
 QY 301 CAACGAGCTGACTTCAATCTCTCGACTGGGGGAGAACTCTGCCAGGGGCAAGGT 360  
 Db 301 CAACGAGCTGACTTCAATCTCTCGACTGGGGGAGAACTCTGCCAGGGGCAAGGT 360  
 QY 361 CAACCTGGAAAGTGGGGAAGCTCAATGACAGGCTGGGCTGACCCAGAACTATGAGG 420  
 Db 361 CAACCTGGAAAGTGGGGAAGCTCAATGACAGGCTGGGCTGACCCAGAACTATGAGG 420  
 QY 421 GTACAGTCACTCTGCTGTTACTTGGTGGCTCAACCGTCAAGCTGAGCTGAACT 480  
 Db 421 GTACAGTCACTCTGCTGTTACTTGGTGGCTCAACCGTCAAGCTGAGCTGAACT 480  
 QY 481 CCGAGCTAGCTGGCCCACTTCTGTACCAAGCTCCAGGGCTGCTGGGAGCAATTCAGG 540  
 Db 481 CCGAGCTAGCTGGCCCACTTCTGTACCAAGCTCCAGGGCTGCTGGGAGCAATTCAGG 540  
 QY 541 TGTATGGGAGCTTGGCTTACCACTGCCCCAGGCTCTGCCAGGAGTCAAGCCGCTG 600  
 Db 541 TGTATGGGAGCTTGGCTTACCACTGCCCCAGGCTCTGCCAGGAGTCAAGCCGCTG 600  
 QY 601 GGGCCCTGGCCCTGCCAGTGAATCTCTCAGAAAGATGGATGACTTCTGGCTGTGAA 660  
 Db 601 GGGCCCTGGCCCTGCCAGTGAATCTCTCAGAAAGATGGATGACTTCTGGCTGTGAA 660  
 QY 661 GGAGCTGACAGCTGGCTATGGCTTACCGGAGGCTTCAACCGCTTAAGAAAGAT 720  
 Db 661 GGAGCTGACAGCTGGCTATGGCTTACCGGAGGCTTCAACCGCTTAAGAAAGAT 720  
 QY 721 GCAGCTCCAGAGCTTCACTCAGCTTGGAGGACATGGTTCTGACCTCTGAC 780  
 Db 721 GCAGCTCCAGAGCTTCACTCAGCTTGGAGGACATGGTTCTGACCTCTGAC 780  
 QY 781 CTTTAAACCCACACTCCAGGCGCAGTCACTGCTGCTT 819  
 Db 781 CTTTAAACCCACACTCCAGGCGCAGTCACTGCTGCTT 819

## RESULT 5

AAH74484

ID AAH74484 standard; DNA; 881 BP.

AC AAH74484;

XX 15-OCT-2001 (first entry)

XX Nucleotide sequence of a human NNT-1 protein.

KW NNT-1; CLF-1; SCNTFRalpha; nervous system; neuron; nervous system;  
 KW neuro-muscular function; tumour; immune system; haematopoietic system;  
 KW reproductive system; liver; skeletal muscle; neurodegenerative disease;  
 KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;  
 KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;  
 KW blastocyst implantation; thrombosis; retinal disease;  
 KW retinal pigmentosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 174..851

XX /\*tag= a

XX /product= "NNT-1"

XX WO200155172-A2.

XX 02-AUG-2001.

XX

XX

XX

XX

PF

XX

PR

XX

PA

XX

PI

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

26-JAN-2001; 2001WO-FR00253.

27-JAN-2000; 2000FR-0001035.

12-OCT-2000; 2000FR-0001089.

(FABR ) FABRE MEDICAMENT SA PIERRE.

(INRM ) INSERM INST NAT SANTE &amp; RECH MEDICALE.

Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;

WPI; 2001-488773/53.

P-PSDB; AAG63543.

A complex comprising a NNT-1 protein and a CLF-1 and/or SCNTFRalpha

protein useful to treat neurodegenerative disease including Parkinson's

and Huntington's, obesity and cancer -

Disclosure; Page 57-58; 67pp; French.

The present sequence encodes a human NNT-1 protein. The specification

describes a complex comprising a NNT-1 protein and a CLF-1 and/or

SCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate

activity of the SCNTFRalpha/gp130/LIFRbeta receptor complex, or to

induce phosphorylation of the tyrosine of gp130 and LIFRbeta.

Particularly where cells expressing the receptor complex are in the

central or peripheral nervous system, in neurons implicated in the

neuro-muscular function or in skeletal muscle. The complex or

antibodies are also used to decrease the survival, growth or

proliferation of tumour cells or to facilitate the proliferation and/or

inhibit differentiation of cells stocks. The complex is also used to

modulate activity of the gp130/LIFRbeta receptor or cells expressing

that receptor, particularly those cells implicated in the immune,

haematopoietic, nervous or reproductive system, the liver or skeletal

muscle. Molecules of the invention may be used to prevent or treat

neurodegenerative diseases including amyotrophic lateral sclerosis,

Parkinson's and Huntington's disease, to repair or regenerate nervous

or muscular tissue or to maintain muscular mass in paralysis patients.

They may also be used to treat cancer, obesity and associated diseases,

and to improve fertility, particularly to avoid endometriosis and/or

assist blastocyst implantation, thrombosis, or retinal disease,

particular retinal pigmentosis.

Sequence 881 BP; 158 A; 318 C; 246 G; 159 T; 0 other;

Query Match 82.3%; Score 674; DB 22; Length 881;

Best Local Similarity 92.2%; Pred. No. 1.5e-159;

Matches 721; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

Qy 2 ATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGGCTCTGGGAGAG 61

Db 82 ATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGGCTCTGGGAGAG 141

Qy 62 GAGCG 121

Db 142 GAGCG 200

Qy 122 GGGATGTTAGCTTGCCTATGACGCTGCTGCGGAGCTTCCCTCCGAGTCCGAGCTTTAAT 181

Db 201 GGGATGTTAGCTTGCCTATGACGCTGCTGCGGAGCTTCCCTCCGAGTCCGAGCTTTAAT 260

Qy 182 CGCAGAGATCCAGGCG 241

Db 261 CGCAGAGATCCAGGCG 320

Qy 242 CTGGAGCATCACTCCGAGCTTAGCTGGGACCTACTGAACTACTGGGGGCGCCCTTTC 301

Db 321 CTGGAGCATCACTCCGAGCTTAGCTGGGACCTACTGAACTACTGGGGGCGCCCTTTC 380

Qy 302 AACGAGCTTGAATCTCAATCTCTCCGAGTGGGGGAGAACTCTGCCAGGGGCAAGGT 361

Db 381 AACGAGCTTGAATCTCAATCTCTCCGAGTGGGGGAGAACTCTGCCAGGGGCAAGGT 440

Qy 362 AACTTGAAGTGGGCGGAGCGCTCAATGACAGGCTGCGGCTGAGCCAGAACTATGAGCG 421

Db 441 GACTTGGAGGTGTGGCGAAGCCTCAATGACAAACTCGGGCTGACCCAGAACTACGAGGCC 500  
 Qy 422 TACAGTCACCTCTGTGTACTTGGCTGCTCAACCGTCAGGCTGCCAGCTGAATC 481  
 Db 501 TACAGCCACCTTCTGTGTACTTGGCTGCTCAACCGTCAGGCTGCCAGCTGAATC 560  
 Qy 482 CGACGTAGCTGGCCCACTTCTGTACAGCCTCCAGGCTGCTGGGCGACATTCAGGT 541  
 Db 561 CGCGCAGCTGGCCCACTTCTGTACAGCCTCCAGGCTGCTGGGCGACATTCAGGT 620  
 Qy 542 GTATGGGAGCTGGCTTACCACTGCCCGCTGCTGAGGCTGCTGAGGCTGCTGAG 601  
 Db 621 GTATGGGAGCTGGCTTACCACTGCCCGCTGCTGAGGCTGCTGAGGCTGCTGAG 680  
 Qy 602 GCCCTGGGCTGCCACAGTGTCTCTCCAGAGATGATGATGATGATGATGATGATG 661  
 Db 681 ACTCTGGGCTGCCACAGTGTCTCTCCAGAGATGATGATGATGATGATGATGATG 740  
 Qy 662 GAGCTGCAGACCTGGCTATGCGCTTACGCAAGGCTTCAACCGGCTTAAGAAAGATG 721  
 Db 741 GAGCTGCAGACCTGGCTATGCGCTTACGCAAGGCTTCAACCGGCTTAAGAAAGATG 800  
 Qy 722 CAGCTCCAGACCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781  
 Db 801 CAGCTCCAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 860  
 Qy 782 CT 783  
 Db 861 TT 862

RESULT 6  
 AAA88546  
 ID AAA88546 standard; DNA; 1790 BP.  
 XX  
 AC AAA88546;  
 XX  
 XX  
 DT 22-JAN-2001 (first entry)  
 XX  
 DE Human interleukin-B60 (IL-B60) gene.  
 XX  
 KW Interleukin-B60; IL-B60; human; cytokine; chromosome 11;  
 KW cytokine-like factor-1; haematopoietic; inflammation;  
 KW antiinflammatory; autoimmune disease; therapy; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 162..809  
 FT /\*tag= a  
 FT sig\_peptide 162..212  
 FT /\*tag= b  
 FT mat\_peptide 213..806  
 FT /\*tag= c  
 XX  
 PN WO200053631-A1.  
 XX  
 XX  
 PD 14-SEP-2000.  
 XX  
 XX 09-MAR-2000; 2000WO-US06182.  
 XX  
 XX 11-MAR-1999; 99US-0267901.  
 XX  
 XX (SCHE ) SCHERING CORP.  
 XX  
 XX Oppmann B, Timans JC, Kastelein RA, Bazan JF;  
 XX  
 XX WPI; 2000-587426/55.  
 XX P-PSDB; AAB19686.  
 XX  
 XX Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes,  
 XX polypeptides, and nucleic acids, useful in research, diagnosis and for

treating inflammatory and autoimmune disorders -

Claim 17; Page 15-16; 97pp; English.

The present sequence is that of DNA encoding human interleukin-B60 (IL-B60, see AAB19586), a novel, small soluble cytokine-like protein of 198 amino acids that exhibits structural motifs characteristic of a member of the long-chain cytokines, and which shows homology to granulocyte colony stimulating factor and interleukin-6. IL-B60 may have either stimulatory or inhibitory effects on haematopoietic cells, including e.g. lymphoid cells, such as T-cells, B-cells, natural killer cells, macrophages, dendritic cells, haematopoietic progenitors, etc. Methods are provided for modulating the physiology or development of a cell or tissue culture cells by contacting the cell with an agonist or antagonist of IL-B60 or an agonist of antagonist of a complex of mature IL-B60 and its partner, cytokine-like factor-1 (CLF-1, see AAB19588). The IL-B60/CLF-1 cytokine serves as a key physiological factor in motor neuron development and regeneration. IL-B60, its agonists and antagonists may be used to treat inflammatory or autoimmune disorders and also for drug screening. The IL60B gene maps to human chromosome 11.

Sequence 1790 BP; 381 A; 560 C; 474 G; 375 T; 0 other;

Query Match 82.3%; Score 674; DB 21; Length 1790;

Best Local Similarity 92.2%; Pred. No. 1.8e-159;

Matches 721; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

Qy 2 ATTATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGGAGAG 61  
 Db 40 ATTATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGGAGAG 99  
 Qy 62 GAGCG 121  
 Db 100 GAGCG 158  
 Qy 122 GGGATTTAGCTTGCCTATGCGAGGTGTGTGCGACCTCCCTCGAGTCCAGCTTAAAT 181  
 Db 159 GGGATTTAGCTTGCCTATGCGAGGTGTGTGCGACCTCCCTCGAGTCCAGCTTCAAT 218  
 Qy 182 CGCAGAGGAGATCCAGGCGCTGCGCCCTCCATCCAGAAAACCTATGACCTCACCGGTAC 241  
 Db 219 CGCAGAGGAGATCCAGGCGCTGCGCCCTCCATCCAGAAAACCTATGACCTCACCGGTAC 278  
 Qy 242 CTGGAGCATCACTCCGCGAGCTTAGTGGGACCTACTGAACTACCTGGGCGCCCTTTC 301  
 Db 279 CTGGAGCATCACTCCGCGAGCTTAGTGGGACCTACTGAACTACCTGGGCGCCCTTTC 338  
 Qy 302 AACGAGCTGACTTCAATCCTCTCGACTGGGGGCGAGAACTCTGCCAGGGCCACCGTTC 361  
 Db 339 AACGAGCTGACTTCAATCCTCTCGACTGGGGGCGAGACTCTGCCAGGGCCACCTGTT 398  
 Qy 362 AACTTGGAGTGTGGGAGAGCTCAATGACAGGCTGGGGCTGAGCCAGAACTATGAGGCG 421  
 Db 399 GACTTGGAGTGTGGGAGAGCTCAATGACAGGCTGGGGCTGAGCCAGAACTATGAGGCG 458  
 Qy 422 TACAGTCACCTCTGTGTACTTGTGCGTGGCTCAACCGTCAAGTCCAGAGTCAATC 481  
 Db 459 TACAGTCACCTCTGTGTACTTGTGCGTGGCTCAACCGTCAAGTCCAGAGTCAATC 518  
 Qy 482 CGAGGTAGCTGGCGGCTTCTGTACAGGCTCCAGGGCGCTGCTGGGCGAGCATTCAGGT 541  
 Db 519 CGCGCGAGCTGGCGGCTTCTGTACAGGCTCCAGGGCGCTGCTGGGCGAGCATTCAGGT 578  
 Qy 542 GTCATGGCGAGCTTGGCTTACCCACTGCGCGCTGCTGGGCGAGCATTCAGGT 601  
 Db 579 GTCATGGCGAGCTTGGCTTACCCACTGCGCGCTGCTGGGCGAGCATTCAGGT 638  
 Qy 602 GCCCTGGCGGCTGGCGGCTTCTGTACAGGCTCCAGAGATGAGTCTTCTGGCTGCTGAAG 661  
 Db 639 ACTCTGGCGGCTGGCGGCTTCTGTACAGGCTCCAGAGATGAGTCTTCTGGCTGCTGAAG 698



RESULT 8  
AAV22652  
ID AAV22652 standard; cDNA; 797 BP.  
XX  
AC AAV22652;  
XX  
DT 13-JUL-1998 (first entry)  
XX  
XX cDNA encoding human neurotrophic factor NNT-1.  
XX  
XX Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;  
KW treatment; neurological disease; degeneration; Parkinson's disease;  
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 90..767  
FT /\*tag= a  
FT sig\_peptide 90..170  
FT /\*tag= b  
FT mat\_peptide 171..764  
FT /\*tag= c  
XX  
XX US57411772-A.  
PN  
PD 21-APR-1998.  
XX  
XX 03-FEB-1997; 97US-0792019.  
PF  
XX 03-FEB-1997; 97US-0792019.  
PR  
XX (AMGE-) AMGEN INC.  
PA  
PI Chang M;  
XX  
XX WPI; 1998-260526/23.  
DR P-PSDB; AAW56141.  
XX  
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -  
PT useful for stimulating growth of motor and sympathetic neurons  
XX  
XX Disclosure; Fig 1; 41pp; English.  
XX  
XX The present sequence encodes a human neurotrophic factor, designated  
CC NNT-1, which is capable of stimulating growth of motor or sympathetic  
CC neurons. The NNT-1 protein is useful in the treatment of neurological  
CC diseases characterised by the degeneration and death of particular  
CC classes of neurons. These diseases specifically include Parkinson's  
CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,  
CC stroke and various degenerative disorders affecting vision.  
XX  
XX Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;  
SQ  
Query Match 81.7%; Score 669.4; DB 19; Length 797;  
Best Local Similarity 92.0%; Pred. No. 2.1e-158;  
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
5 ATTTAAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGGCAGGCTCTGGAGAGGAG 64  
1 ATTTAAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGGCAGGCTCTGGAGAGGAG 60  
65 CG 124  
61 CGCGACCG 119  
125 ATGTTAGCTTGCTATGCAAGGCTGTGGACCTCCCTGCACTGCGAGCTCTTAATCGC 184  
120, ATGTTAGCTGCTGTGCAAGGCTGTGGACCTCCCTGCACTGCGAGCTCTCAATCGC 179  
185 ACAGGAGATCCAGGCG 244  
180 ACAGGAGATCCAGGCG 239

QY 245 GAGATCAACTTCGCGAGCTTAGCTGGACCTTACCTGAACCTACCTGGGGCCCCCTTTCAAC 304  
DB 240 GAGACCAACTTCGCGAGCTTAGCTGGACCTTATCTGAACCTACCTGGGGCCCCCTTTCAAC 299  
QY 305 GAGCCTGACTTCAATCTCTCTGAGCTGGGGGCGAGAACTCTGCCAGGCGCACGCTCAAC 364  
DB 300 GAGCCAGACTTCAACCTCTCCCGCTGGGGGCGAGAGACTCTGCCAGGCGCACGCTGTGAC 359  
QY 365 TTGGAAGTGTGGCGAAGCCTCAATGACAGGCTGGGGCTGAGCCAGAACTATGAGGCGTAC 424  
DB 360 TTGAGAGTGTGGCGAAGCCTCAATGACAACTGGGGCTGAGCCAGAACTATGAGGCGTAC 419  
QY 425 AGTCACCTCTCTGTGTACTTCTGCTGGCTCAACCGCTCAGGCTGCCACAGCTGAATCCGA 484  
DB 420 AGCCACCTTCTGTGTACTTCTGCTGGCTCAACCGCTCAGGCTGCCACAGCTGTGAGCTGCGC 479  
QY 485 CGTAGCCTGGCCCACTTCTGTACAGGCTCCAGGCGCTGCGGCGAGCAATGAGGCTGTG 544  
DB 480 CGCAGCCTGGCCCACTTCTGTACAGGCTCCAGGCGCTGCTGGGCGAGCAATGAGGCGTAC 539  
QY 545 ATGGCGAGCTGGCTACCCACTGCGCCAGCCTCTGCCAGGAGCTGAGCAGCCTGGGCC 604  
DB 540 ATGGCAGCTCTGGGCTACCCACTGCGCCAGCCTGCGCTGCCAGCTGAAACCCACTTGGACT 599  
QY 605 CTTGGCCCTGGCCCACTGAGTACTTCTCCAGAAAGATGGAGCTTCTGGCTGCTGAAGGAG 664  
DB 600 CTTGGCCCTGGCCCACTGAGTACTTCTCCAGAAAGATGGAGCTTCTGGCTGCTGAAGGAG 659  
QY 665 CTGACAGCTGGCTATGGGTTTCCAGCAGGAGCTTCAACCGGCTTAAGAAAGAGATGCGAG 724  
DB 660 CTGACAGCTGGCTATGGGTTTCCAGCAGGAGCTTCAACCGGCTTAAGAAAGAGATGCGAG 719  
QY 725 CTTCCAGCAGCTTCACTGCTGAGTCACTTGGAGGCACTGGTCTTCTGACCTCTGACCTT 783  
DB 720 CTTCCAGCAGCTGAGTCACTTGGAGGCTTGGAGGCTTCTGAGCTTCTGACCTTCTGACCTT 778  
RESULT 9  
AAA39481  
ID AAA39481 standard; cDNA; 797 BP.  
XX  
AC AAA39481;  
XX  
DT 24-AUG-2000 (first entry)  
XX  
XX Human NNT-1 cDNA.  
XX  
XX NNT-1; human; neurotrophic factor; neurotropic; neuroprotective; treatment;  
CC anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;  
CC nervous system degeneration; Alzheimer's disease; Parkinson's disease;  
CC amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
CC Huntington's disease; peripheral neuropathy; neural retina degeneration;  
CC retinopathy; immune disorder; hematopoietic disorder; ss.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 90..767  
FT /\*tag= a  
FT /product= "NNT-1"  
XX  
XX US6054294-A.  
XX  
XX 25-APR-2000.  
XX  
XX 12-DEC-1997; 97US-0988819.  
XX  
XX 03-FEB-1997; 97US-0792019.  
XX  
XX (AMGE-) AMGEN INC.  
XX  
XX Chang M;  
PI





100







```

RESULT 14
AAD04201
ID AAD04201 standard; cDNA; 729 BP.
XX
AC AAD04201;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human cardiotrophin-like cytokine (CLC) cDNA.
XX
KW Human; biologically active complex; haemopoietin receptor; NR6;
KW cardiotrophin-like cytokine; CLC; therapy; prophylaxis; proliferation;
KW differentiation; cell survival; neurotrophic activity; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 11..679
FT /tag= a
FT /product= "Human cardiotrophin-like cytokine (CLC)
FT /note= "CDS does not include stop codon"
FT /partial
FT sig_peptide 11..91
FT mat_peptide 92..679
FT /tag= b
FT /tag= c
FT /product= "Human mature cardiotrophin-like cytokine
FT (CLC) protein"
XX
PN WO200127157-A1.
XX
PD 19-APR-2001.
XX
PF 06-OCT-2000; 2000WO-AU01216.
XX
PR 08-OCT-1999; 99AU-0003327.
XX
PR 12-MAY-2000; 2000AU-0007489.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;
PI Nakata Y, Hasegawa M;
XX
WPI; 2001-281978/29.
XX
P-PSDB; AAE00828.
XX
PT New biologically active complex comprising NR6 and
PT cardiotrophin-like-cytokine, for facilitating proliferation,
PT differentiation and/or survival of a cell -
XX
PS Claim 31; Page 112-114; 123pp; English.
XX
CC The present invention relates to a biologically active complex comprising
CC a haemopoietin receptor, NR6 and cardiotrophin-like cytokine (CLC).
CC The complex is useful in the manufacture of a medicament for the
CC treatment and/or prophylaxis of a subject, as it is involved in
CC facilitating proliferation, differentiation and/or survival of a cell.
CC The complex or its components have neurotrophic activity. The present
CC sequence is human cardiotrophin-like cytokine (CLC) cDNA.
XX
SQ Sequence 729 BP; 132 A; 261 C; 196 G; 140 T; 0 other;
Query Match 73.3%; Score 600.2; DB 22; Length 729;
Best Local Similarity 91.6%; Pred. No. 4.7e-141;
Matches 635; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 91 CCCATGACCTCCGACGAGGGGACTCGTGGGGATGTTAGCTTGCTATGACGGTGTCT 150
DB 7 CCCATGACCTCCGACGAGGGGACTCGTGGGGATGTTAGCTTGCTATGACGGTGTCT 66
QY 151 GTGGCAGCTCCCTGCAGTGCAGCTTTAATCGACAGGAGATCCAGGCCCTGGCCCTC 210

```

```

Db 67 CTGGCAGCTCCCTGCAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTC 126
QY 211 CATCCAGAAAACTATGACCTACCCGCTACCTGGAGCATCACTCCGAGCTTAGCTGG 270
Db 127 CATCCAGAAAACTATGACCTACCCGCTACCTGGAGCATCACTCCGAGCTTAGCTGG 186
QY 271 GACCTACCTGAACCTACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCTCTCCTGACT 330
Db 187 GACCTATCTGAACCTACCTGGGGCCCCCTTTCAACGAGCAGACTTCAACCTCCCGCCT 246
QY 331 GGGGGCAGAAAACTCTGCCAGGGGCCAGCTCACTTGAAGTGTGGGAAAGCTCAATGA 390
Db 247 GGGGGCAGAGACTCTGCCAGGGGCCACTGTGTGACTTGGAGGTGGGCGAAGCCTCAATGA 306
QY 391 CAGGCTGCGGCTGACCCAGAACTATGAGGCGTACAGTCACTCTGTGTTTACTTGGCTGG 450
Db 307 CAAACTGCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACTTGGCTGG 366
QY 451 CCTCAACCGCTCAGGCTGCCACAGCTGAACCTCGACCTAGCTGGCCCACTTCTGTATCCAG 510
Db 367 CTTCAACCGCTCAGGCTGCCACTGCTGAGCTGGCCGAGCTGGCCCACTTCTGTGACCAAG 426
QY 511 CCTCAGGGCCTGCTGGGCGAGCAATTGCAGGTGTCAAGTGTGATGGCGAGCTTGGCTACCCACTGCC 570
Db 427 CCTCAGGGCCTGCTGGGCGAGCAATTGCAGGTGTGATGGCGAGCTTGGCTACCCACTGCC 486
QY 571 CAGGCTTGGCAGGGAGCTGAGCCAGCTTGGGCCCCCTGGCCCTGCCACAGTACTTCTT 630
Db 487 CCAGCGCTGCTGGGAGCTGAACCCCACTTGGACTCTCTGCCCCCTGCCACAGTACTTCTT 546
QY 631 CCAGAGATGATGATGACTTCTGGCTGTGAGAGAGATGCAGAGCTGAGAGCTTGGCTATGGCGTTT 690
Db 547 CCAGAGATGATGAGACTTCTGGCTGTGAGAGAGATGCAGAGCTTGGCTATGGCGTTT 606
QY 691 CAAGGACTTCAACCGGCTTAAAGAGAGATGCAGAGCTTCCAGAGCTTCCAGTCAACCTGCA 750
Db 607 CAAGGACTTCAACCGGCTTCAAGAGAGATGCAGAGCTTCCAGAGCTTCCAGTCAACCTGCA 666
QY 751 CTTGGAGGCACATGTTTCTGACCTTCTGACCTTCTGACCTT 783
Db 667 CTTGGGGCTCATGGCTTCTGACCTTCTGACCTT 699
RESULT 15
AAK51548
ID AAK51548 standard; cDNA; 1008 BP.
XX
AC AAK51548;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 93.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.

```

QY	640	GGATGACTTCTGGCTGTCTGAAGAGAGCTGCAGACCTTGGCTATGGCGTTT	CAGCCCAAGGACATT	699
Db	836	GGACGACTTCTGGCTGTCTGAAGAGAGCTGCAGACCTTGGCTGTGGCGCT	CGGCCCAAGGACATT	895
QY	700	CAACCGCGCTTAAGAAGAGATGCAGCCTCCAGCAGCTTCAGTCACCCCTGC	ACTTTGGAGGC	759
Db	896	CAACCGCGCTCAAGAAGAAGATGCAGCCTCCAGCAGCTCGAGTCACCCCT	GCACCTGGGGGC	955
QY	760	ACATGGTTTCTGACCTCTGACCCT		783
Db	956	TCATGGCTTCTGACTTCTTGACCTT		979

```
Query Match      71.2%; Score 583.2; DB 22; Length 1008;
Best Local Similarity 90.8%; Pred. No. 9.5e-137;
Matches 621: Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```

QY	100	CTTCGAGCAGGGGACTCGTGGGGGATGTACTTCGCTATGCATGACGGTCTGGGCACCT	150
Db	296	CCTGCCACAGGGGACTCGTGGGGGATGTTACGCTGCCCTGTGCACGTCTCTGGCACCT	355
QY	160	CCCTGCAGTGGCAGCTCTTAATCGCACAGGAGATCCAGGCCTCGCCCCCTCAATCCAGAA	219
Db	356	CCCTGCAGTGGCAGCTCTCAATCGCACAGGGACCCAGGGCCTGGCCCCCTCATCCAGAA	415
QY	220	AACCTATGACCTCACCGGTACCTTGGAGCATCAACTCCGACAGCTTAGCTGGGACCTACT	279
Db	416	AACCTATGACCCACCGCTTACCTGGAGCACCAACTCCGACGCTTGGCTTGGGACCTATCT	475
QY	280	GAATACCTGGGGCCCCCTTTCAACGAGCGCTGACTTTCATCTCTCTCGACTGGGGCAGA	339
Db	476	GAATACCTGGGGCCCCCTTTCAACGAGCGAGACTTCAACCTCCCGCCTGGGGGCAGA	535
QY	340	AACCTGCGCCAGGGGCACCGGTCAACTTGGAGATGTGGCGAAGCCTCAATGACAGGCTCGG	399
Db	536	GACTCTGCCAGGGGCCACTTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAACTGCG	595
QY	400	GCTGACCCAGAACTATGAGCGGTACAGTCACTCTGTGTACTTTCGCTGGGCTCAACCG	459
Db	596	GCTGACCCAGAACTATGAGGCTTACAGCACCTTCTGTGTACTTTCGCTGGGCTCAACCG	655
QY	460	TCAGGCTGCCACAGCTGAACTTCGACGTAGCTTGGGCCCATTTCTGTACAGGCTCCAGG	519
Db	656	TCAGGCTGCCACTGCTGAGCTTGGCGCGACCTTGGGCCCATTTCTGTACAGGCTCCAGG	715
QY	520	CCTGCTGGGACGATTCGAGGTGCATGGCGAGCTTGGCTACCCACTGCCCGACGCTCT	579
Db	716	CCTGCTGGGACGATTTGGGGCGTTCATGGCAGTCTTGGGCTACCCACTGCCCGACGCTCT	775
QY	580	GCCAGGGACTGAGCCAGCGCTTGGGGCCCTTGGCCCTTGCACAGTGAATTTCTCCAGAGAT	639
Db	776	GCCTTGGACTGAACCCCACTTGGACTCTTGGCCCTTGGCCCGCCAGTGAATTTCCACAGAAT	835

db 1 TATTATAAGCTTCGCCGGAGCCGGGCTCGCCCTCCCACTCCGCCAGCCTCTGGGAGA 60



```
FILE REFERENCE: A-695
CURRENT APPLICATION NUMBER: US/09/931,704
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: US 60/226,436
PRIOR FILING DATE: 2000-08-18
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patent in version 3.1
SEQ ID NO 3
LENGTH: 5087
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (137)..(138)
OTHER INFORMATION: product = "INTERVENING UNSEQUENCED REGION OF >1KB"
US-09-931-704-3

Query Match          52.5%; Score 429.6; DB 10; Length 5087;
Best Local Similarity 90.4%; Pred. No. 5.5e-105;
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 276 ACCTGAAGTACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCCTCCTCGACTGGGG 335
Db 3363 AGCTGAAGTACCTGGGGCCCCCTTTCAACGAGCAGACTTCAACCTTCCCGCCTGGGG 3422
QY 336 CAGAAACTCTGCCAGGCCCAAGCTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGGC 395
Db 3423 CAGAGACTCTGCCAGGCCCACTGTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 3482
QY 396 TGGGGTGAGCCAGCAATATGAGCGGTACAGTCACTCCTGTGTACTTGGGTGGCCTCA 455
Db 3483 TGGGGTGAGCCAGCAATATGAGCGGTACAGTCACTCCTGTGTACTTGGGTGGCCTCA 3542
QY 456 ACCGTGAGGCTGCCAGAGCTCAAGCTCGAGCTAGCCTGGCCCACTTCTGTACCAAGCTCC 515
Db 3543 ACCGTGAGGCTGCCAGAGCTCAAGCTCGAGCTAGCCTGGCCCACTTCTGTACCAAGCTCC 3602
QY 516 AGGCGCTGCTGGGAGCATTGACAGTGTGATGGGAGCGCTTGGTACCCACTGCCCCAGC 575
Db 3603 AGGCGCTGCTGGGAGCATTGACAGTGTGATGGGAGCGCTTGGTACCCACTGCCCCAGC 3662
QY 576 CTGTGCAGGAGTACGAGCAGCTGGGCCCTGGCCCTGCCAGTGCAGTACTTCTCTCAGA 635
Db 3663 CGTGTGCTGGGAGTGAACCCACTTGGACTCTCTGGCCCTGCCAGTGCAGTACTTCTCTCAGA 3722
QY 636 AGATGGATGACTTCTGCTGCTGAAGAGCTGACAGCTGCTGCTATGGCTTTCAGCCAAAG 695
Db 3723 AGATGGAGACTTCTGCTGCTGAAGAGCTGACAGCTGCTGCTATGGCTTTCAGCCAAAG 3782
QY 696 ACTTCAACCGGCTTGAAGAAGATGACAGCTTCCAGCAGCTTTCAGTCACTTGCACCTTGG 755
Db 3783 ACTTCAACCGGCTTGAAGAAGATGACAGCTTCCAGCAGCTTTCAGTCACTTGCACCTTGG 3842
QY 756 AGGCACATGTTTCTGACCTTGCACCT 783
Db 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870

RESULT 4
US-09-864-761-23175/c
Sequence 23175, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aesomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
```

```
FILE FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 23175
LENGTH: 495
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005849.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
OTHER INFORMATION: SWISSPROT HIT: O63086, EVALUATION 8.00e-03
OTHER INFORMATION: NT HIT: g11439486, EVALUATION 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: A1752561.1, EVALUATION 0.00e+00
US-09-864-761-23175
```

```
Query Match          51.3%; Score 419.8; DB 10; Length 495;
Best Local Similarity 90.5%; Pred. No. 1e-102;
Matches 448; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 278 CTGAATACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCCTCTCGACTGGGGCA 337
Db 495 CTGAATACCTGGGGCCCCCTTTCAACGAGCAGACTTCAACCTTCCCGCCTGGGGCA 436
QY 338 GAAACTCTGCCAGGCCCAAGCTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGGCTG 397
Db 435 AAGACTCTGCCAGGCCCACTGTGTGAGTGTGGAGGTGTGGCGAAGCCTCAATGACAACTG 376
QY 398 CGGCTGACCCAGCAACTATGAGCGCTCAGTCACTCCTCTGTGTACTTGCCTGGGCTCAAC 457
Db 375 CGGCTGACCCAGCAACTAGAGGCCCTCAGCCACCTTCTGTGTACTTGCCTGGGCTCAAC 316
QY 458 CGTCAGGCTGCCACAGCTGAACCTCCGAGCTAGCTGCGCCCACTTCTGTACCAAGCTCCAG 517
```

Db 315 CGTCAGGCTGCCACTGCTAGCTGCGCGCAGCCTGGCCCACTTCTGCACGAGCTCCAG 256  
QY 518 GCCTGCTGGGAGCATTGCGAGTGTATGCGACGCTTGGCTACCCAGCTCCAGCCT 577  
Db 255 GGCCTGCTGGGAGCATTGCGCGCTCATGCGAGCTTGGGCTACCCAGCTGCCAGCG 196  
QY 578 CTGCCAGGAGCTGAGCAGCCTGGGCGCCCTGGCCCTGCCACAGTACTTCTCCAGAAG 637  
Db 195 CTGCTGGAGCTGAACCCACTTGGATCTCTGGCCCTGCCACAGTACTTCTCCAGAAG 136  
QY 638 ATGATGACTTCTGGCTGTGAAGAGCTGCAGACCTGGCTATGGGCTTTCAGCCAAAGGAC 697  
Db 135 ATGAGCAGCTTCTGGCTGTGAAGAGCTGCAGACCTGGCTGTGGGCTCGGCCAAGGAC 76  
QY 698 TTCAACCGGCTTAAGAGAGATGCGACCTCCAGCAGCTTTCAGTCACTCCCTGCAGTTGAG 757  
Db 75 TTCAACCGGCTCAAGAGAGATGCGACCTCCAGCAGCTTTCAGTCACTCCCTGCAGTTGAG 16  
QY 758 GCACATGGTTTCTGA 772  
Db 15 GTCATGGCTTCTGA 1

RESULT 5  
US-09-864-761-6462/c  
; Sequence 6462, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aomic-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 6462  
; LENGTH: 492  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005849.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5  
; US-09-864-761-6462

Query Match 36.5%; Score 299.2; DB 10; Length 492;  
Best Local Similarity 90.6%; Pred. No. 1.3e-70;  
Matches 319; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 432 TCCTGTGTTACTTGGTGGCTCAACCGTCAAGGCTGCCAGCTGAACCTCCAGCGTAGCC 491  
Db 492 TTCTGTGTTACTTGGTGGCTCAACCGTCAAGGCTGCCAGCTGAACCTCCAGCGTAGCC 433  
QY 492 TGGCCCACTTCTGTACCAAGCTCCAGGCGCTGCTGGGAGCAGATTGACGTGTCTATGGCGA 551  
Db 432 TGGCCCACTTCTGTACCAAGCTCCAGGCGCTGCTGGGAGCAGATTGACGTGTCTATGGCGA 373  
QY 552 CGCTTGGCTACCCACTGCCCCAGCCTCTGCCAGGAGCTGAGCAGCTGGGCGCCCTGGGCC 611  
Db 372 CTCTGGGCTACCCACTGCCCCAGCCTGCTGGGAGCAGATTGACGTGTCTATGGCGC 313  
QY 612 CTGCCCCAGTGACTTCTCCAGAAAGATGGATGACTTCTGGCTGTGAAAGAGCTGCAGA 671  
Db 312 CTGCCCCAGTGACTTCTCCAGAAAGATGGATGACTTCTGGCTGTGAAAGAGCTGCAGA 253  
QY 672 CTTGGCTATGGCGTTCAGCCAGGAGCTTCAACCGGCTTAAGAGAGATGACGCTCCAG 731  
Db 252 CTTGGCTATGGCGTTCAGCCAGGAGCTTCAACCGGCTTAAGAGAGATGACGCTCCAG 193  
QY 732 CAGCTTCAGTCACCTGCAGCTTGAGGAGCAGATGTTCTGACCTCTGACCT 783  
Db 192 CAGCTTCAGTCACCTGCAGCTTGAGGAGCTTCAACCGGCTTAAGAGAGATGACGCTTCTGACCTT 141

RESULT 6  
US-09-765-231A-75  
; Sequence 75, Application US/09765231A  
; Patent No. US20020119452A1  
; GENERAL INFORMATION:  
; APPLICANT: Searle/Monsanto  
; APPLICANT: Phippard, Deborah  
; APPLICANT: Vasanthakumari, Geetha  
; APPLICANT: Dotson, Stanton  
; APPLICANT: Ma, Xiao-Jun  
; TITLE OF INVENTION: Osteoarthritis tissue-derived nucleic acids, polypeptides,  
; FILE REFERENCE: SO-3221 PR  
; CURRENT APPLICATION NUMBER: US/09/765,231A  
; NUMBER OF SEQ ID NOS: 82  
; SEQ ID NO 75  
; LENGTH: 283  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-765-231A-75

Query Match 28.8%; Score 236; DB 10; Length 283;  
Best Local Similarity 94.0%; Pred. No. 7.3e-54;  
Matches 267; Conservative 0; Mismatches 15; Indels 2; Gaps 2;



QY 453 TCACCGCTCAGGCTGCCACAGCT 475  
Db 8772 TTAGCCAGATGCCAGCCCCCGTT 8794

RESULT 10  
US-09-764-877-3474  
; Sequence 3474, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3474  
; LENGTH: 19172  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-877-3474

Query Match 4.7%; Score 38.2; DB 10; Length 19172;  
Best Local Similarity 58.3%; Pred. No. 1.4;  
Matches 67; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 10 AGCTTCGCGGAGCGCGCTCCCTCCACATCCCGACGCTTGGGAGAGAGCGCG 69  
Db 7171 AGCTTCGCGGAGCGCGCTCCCTCCACATCCCGACGCTTGGGAGAGAGCGCG 69

QY 70 CCGGCG 124  
Db 7231 CCGTGTCCACCG 7285

RESULT 11  
US-09-822-849A-270  
; Sequence 270, Application US/09822849A  
; Patent No. US20020045170A1  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fectel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalak  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6403  
; CURRENT APPLICATION NUMBER: US/09/822,849A  
; CURRENT FILING DATE: 2001-09-04  
; PRIOR APPLICATION NUMBER: 60/195,582  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 598  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 270  
; LENGTH: 2149  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-849A-270

Query Match 4.6%; Score 37.4; DB 10; Length 2149;  
Best Local Similarity 50.3%; Pred. No. 1.1;  
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 514 CCAGGCGCTCTGGGAGCATTGAGGTTGATGCGGAGCGTGGCTACCCACTGCCCA 573  
Db 1224 CCGTGGCTTCGGGAGCTGGATTGACAGTATCTCGGCTTCTCAAGGTCCTGCACAGTT 1283

QY 574 GCCTCTGCCAGGAGCTGAGCCCTGGGCGCCCTGGCCCTGCCACAGTACTTCTCTCA 633  
Db 1284 GCTTGTGATGTCCCTGCCTTGGCTTGGCTTCTCTGCGCTTCTCTCATCACCAGCGCA 1343

QY 634 GAAGATGGATGACTTCTGGCTGTGAGGAGCTGCAGACCTGGCTATGGCGTTACGCCAA 693  
Db 1344 TGGGCTGCAGGAGCGCGCGGGTGGAGAGCTGCAGAACCGCATCGCCAGCTGCTGAA 1403

QY 694 GGA 696  
Db 1404 GGA 1406

RESULT 12  
US-09-853-386-89  
; Sequence 89, Application US/09853386  
; Patent No. US20020049151A1  
; GENERAL INFORMATION:  
; APPLICANT: Murphy, Evelyn  
; APPLICANT: Bresnihan, Barry  
; APPLICANT: Conneely, Orla  
; APPLICANT: Fitzgerald, Oliver  
; TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NURR  
; FILE REFERENCE: P01972US1  
; CURRENT APPLICATION NUMBER: US/09/853,386  
; CURRENT FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/203645  
; PRIOR FILING DATE: 2000-05-12  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 89  
; LENGTH: 2481  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-09-853-386-89

Query Match 4.6%; Score 37.4; DB 10; Length 2481;  
Best Local Similarity 50.3%; Pred. No. 1.2;  
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 514 CCAGGCGCTCTGGGAGCATTGAGGTTGATGCGGAGCGTGGCTACCCACTGCCCA 573  
Db 1537 CCGTGGCTTCGGGAGCTGGATTGACAGTATCTCGGCTTCTCAAGGTCCTGCACAGTT 1596

QY 574 GCCTCTGCCAGGAGCTGAGCCCTGGGCGCCCTGGCCCTGCCACAGTACTTCTCTCA 633  
Db 1597 GCTTGTGATGTCCCTGCCTTGGCTTGGCTTCTCTGCGCTTGTCTCATCACCAGCGCA 1656

QY 634 GAAGATGGATGACTTCTGGCTGTGAGGAGCTGCAGACCTGGCTATGGCGTTACGCCAA 693  
Db 1657 TGGGCTGCAGGAGCGCGCGGTGGAGAGCTGCAGAACCGCATCGCCAGCTGCTGAA 1716

QY 694 GGA 696  
Db 1717 GGA 1719

RESULT 13  
US-10-002-600-45  
; Sequence 45, Application US/10002600  
; Patent No. US20020137077A1  
; GENERAL INFORMATION:  
; APPLICANT: Hopkins, Christopher M.  
; APPLICANT: Peterson, David P.  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Hawkins, Phillip R.  
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS  
; FILE REFERENCE: PA-0042 US  
; CURRENT APPLICATION NUMBER: US/10/002,600  
; CURRENT FILING DATE: 2001-10-25  
; PRIOR APPLICATION NUMBER: 60/243,521



```
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 45
; LENGTH: 2626
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Template ID: 1134866.36
US-10-002-600-45

Query Match      4.6%; Score 37.4; DB 12; Length 2626;
Best Local Similarity 50.3%; Pred. No. 1.2;
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 514 CCAGGGCTGCTGGGACGATTGCAGGTGTGATGCGAGCGTTGGCTACCCACTGCCCA 573
Db 1695 CCGTGGCTTCGGGACTGGATGACAGTATCTCGCCCTTCTCAAGGTCCCTGCACAGCTT 1754
QY 574 GCCTCTGCCAGGACTGAGCCAGCGCTGGGCGCCCTGGCCCTGCCACAGTGACTTCTCTCCA 633
Db 1755 GCTTGTGATGCTCCCTGCTGCTGCTCTCTGCTGCTTGTCTCTCATCAGCGCGCA 1814
QY 634 GAAGATGATGACTTCTGGCTGTGAAGAGCTGAGACCTGGCTATGCGGTTCAGGCCAA 693
Db 1815 TGGGCTGAGGAGCGCGCGGCTGGAGAGCTGCAGAACCGCATCGCCAGCTGCTCGAA 1874
QY 694 GGA 696
Db 1875 GGA 1877

RESULT 14
US-10-002-600-44
; Sequence 44, Application US/10002600
; Patent No. US2002013707A1
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002.600
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 44
; LENGTH: 2670
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Template ID: 1134866.26
US-10-002-600-44

Query Match      4.6%; Score 37.4; DB 12; Length 2670;
Best Local Similarity 50.3%; Pred. No. 1.2;
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 514 CCAGGGCTGCTGGGACGATTGCAGGTGTGATGCGAGCGTTGGCTACCCACTGCCCA 573
Db 1735 CCGTGGCTTCGGGACTGGATGACAGTATCTCGCCCTTCTCAAGGTCCCTGCACAGCTT 1794
QY 574 GCCTCTGCCAGGACTGAGCCAGCGCTGGGCGCCCTGGCCCTGCCACAGTGACTTCTCTCCA 633
Db 1795 GCTTGTGATGCTCCCTGCTGCGCTGCTCTCTGCTGCTTGTCTCTCATCAGCGCGCA 1854
QY 634 GAAGATGATGACTTCTGGCTGTGAAGAGCTGAGACCTGGCTATGCGGTTCAGGCCAA 693
Db 1875 GGA 1877
```

```
Db 1855 TGGGCTGCAGGAGCGCGGCTGGAGAGCTGCAGAACCGCATTCGCCAGCTGCTCGAA 1914
QY 694 GGA 696
Db 1915 GGA 1917

RESULT 15
US-09-728-952-60/c
; Sequence 60, Application US/09728952
; Patent No. US20020111302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha L.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020111302A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/728,952
; CURRENT FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: pc_fl_genes Version 2.0
; SEQ ID NO 60
; LENGTH: 1232
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (344)..(1015)
US-09-728-952-60

Query Match      4.5%; Score 36.6; DB 10; Length 1232;
Best Local Similarity 54.0%; Pred. No. 1.5;
Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 392 AGGCTGGCGGTGACCCAGAACTATGAGCGCTACAGTCACTCTCTGTGTACTTGGCTGGC 451
Db 985 AGGCTCCTGCATCTCCGTCATCCCATGATTGGAGTGTCCAGTACTGGCCCCCGCAGGCTGGC 926
QY 452 CTCACCGTCAAGGCTGCCACAGCTGAACCTCCGACGTAGCTGGGCCACTTCTGTACACGC 511
Db 925 CTGAGCCTTCAGGAAGCCCTCGCTGAATCCATCTCAGCCTGGCTCTCTGTACTCCCA 866
QY 512 CTCAGAGGCTGCTGGGCA 530
Db 865 GTAGAGACCTGCTGCCCA 847

Search completed: February 2, 2003, 04:00:10
Job time : 119.189 secs
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 26.3918 Seconds  
(without alignments)  
9516.918 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 819

Sequence: 1 tattattaaagcttcgcccgcg.....aggccagtcagctgtgctt 819

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	819	100.0	819	1	US-08-792-019B-4
2	819	100.0	819	3	US-08-988-819-4
3	819	100.0	819	3	US-09-016-534-4
4	669.4	81.7	797	1	US-08-792-019B-1
5	669.4	81.7	797	3	US-08-988-819-1
6	669.4	81.7	797	3	US-09-016-534-1
7	627	76.6	1710	3	US-09-106-182-1
8	429.6	52.5	5087	1	US-08-792-019B-3
9	429.6	52.5	5087	3	US-08-988-819-3
10	429.6	52.5	5087	3	US-09-016-534-3
11	52	6.3	396	3	US-09-106-182-7
12	43.2	5.3	6803	3	US-08-665-259-19
13	43.2	5.3	6803	3	US-08-782-500-19
14	42.4	5.2	1642	2	US-08-665-037-1
15	42.4	5.2	1642	2	US-08-665-067-1
16	42.4	5.2	1642	2	US-08-732-870-1
17	38.4	4.7	68750	3	US-09-335-409-1
18	38.4	4.7	68750	4	US-09-568-102-1
19	38.4	4.7	68750	4	US-09-567-969-1
20	38.4	4.7	68750	4	US-09-568-480-1
21	38.4	4.7	68750	4	US-09-568-486-1
22	38.4	4.7	68750	4	US-09-568-472-1
23	38.4	4.7	68750	4	US-09-567-899-1
24	34.8	4.2	4403765	4	US-09-103-840A-2
25	34.8	4.2	4411529	4	US-09-103-840A-1
26	34.4	4.2	1225	1	US-08-197-496A-1
27	34.4	4.2	1225	1	US-08-275-370-1

Sequence 18, Appli  
Sequence 9, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 5, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 115, App  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 18, Appli  
Sequence 56, Appli  
Sequence 57, Appli  
Sequence 56, Appli  
Sequence 57, Appli  
Sequence 58, Appli

## ALIGNMENTS

RESULT 1  
US-08-792-019B-4  
; Sequence 4, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUTROTROPIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/792,019B  
; FILING DATE: 03-FEB-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 819 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 95..769  
; FEATURE:  
; NAME/KEY: mat peptide  
; LOCATION: 176..769  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 95..175  
; US-08-792-019B-4

Query Match 100.0%; Score 819; DB 1; Length 819;  
Best Local Similarity 100.0%; Pred. No. 8.3e-206; Indels 0; Gaps 0;  
Matches 819; Conservative 0; Mismatches 0;

```
QY 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCCTCTGGGAGA 60
Db 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCCTCTGGGAGA 60
QY 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 GGGGATGTTAGCTTGTGCTATGACGGTGTGTGACCTCTCCCTGCGAGTCCAGCTCTTAA 180
Db 121 GGGGATGTTAGCTTGTGCTATGACGGTGTGTGACCTCTCCCTGCGAGTCCAGCTCTTAA 180
QY 181 TCGCAGAGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 TCGCAGAGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 241 CCTGGAGCATCACTCGGAGCTTAGTGGAGCTTACCTGAACTACCTGGGGCGCGCGCT 300
Db 241 CCTGGAGCATCACTCGGAGCTTAGTGGAGCTTACCTGAACTACCTGGGGCGCGCGCT 300
QY 301 CAACAGAGCTGACTTCAATCCTCTGCTGACCTGGGGCGAGAACTCTGCCAGGGCCACGGT 360
Db 301 CAACAGAGCTGACTTCAATCCTCTGCTGACCTGGGGCGAGAACTCTGCCAGGGCCACGGT 360
QY 361 CAACCTTGAAGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
Db 361 CAACCTTGAAGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
QY 421 GTACAGTCACTTCTGTTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Db 421 GTACAGTCACTTCTGTTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
QY 481 CGGAGCTAGCTGGCGCACTTCTGTATCCAGCTCCAGGCGCTGCTGGCGAGCATTTGCGAG 540
Db 481 CGGAGCTAGCTGGCGCACTTCTGTATCCAGCTCCAGGCGCTGCTGGCGAGCATTTGCGAG 540
QY 541 TGTCTAGGCGAGCTTGGCTACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db 541 TGTCTAGGCGAGCTTGGCTACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
QY 601 GCGCCCTGCGCGCTGCGCGAGTACTTCTCCAGAGAGATGATGATCTTCTGGCTGCTGAA 660
Db 601 GCGCCCTGCGCGCTGCGCGAGTACTTCTCCAGAGAGATGATGATCTTCTGGCTGCTGAA 660
QY 661 GGAGCTGACAGCTGGCTATGGCTTTCAGCCAGGAGCTTCAACCGGCTTAAGAGAGAT 720
Db 661 GGAGCTGACAGCTGGCTATGGCTTTCAGCCAGGAGCTTCAACCGGCTTAAGAGAGAT 720
QY 721 CGAGCTCCAGAGCTTCACTGACCTGCACTTGGAGGACATGCTTCTGACCTCTGAC 780
Db 721 CGAGCTCCAGAGCTTCACTGACCTGCACTTGGAGGACATGCTTCTGACCTCTGAC 780
QY 781 CTTTAAACCCCACTCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 819
Db 781 CTTTAAACCCCACTCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 819
```

## RESULT 2

```
US-08-988-819-4
; Sequence 4, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 819 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..769
; NAME/KEY: mat_peptide
; LOCATION: 176..769
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..175
; US-08-988-819-4
```

Query Match 100.0%; Score 819; DB 3; Length 819;

Best Local Similarity 100.0%; Pred. No. 8.3e-206;

Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCCTCTGGGAGA 60
Db 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCCTCTGGGAGA 60
QY 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 GGGGATGTTAGCTTGTGCTATGACGGTGTGTGACCTCTCCCTGCGAGTCCAGCTCTTAA 180
Db 121 GGGGATGTTAGCTTGTGCTATGACGGTGTGTGACCTCTCCCTGCGAGTCCAGCTCTTAA 180
QY 181 TCGCAGAGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 TCGCAGAGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 241 CCTGGAGCATCACTCGGAGCTTAGTGGAGCTTACCTGAACTACCTGGGGCGCGCGCT 300
Db 241 CCTGGAGCATCACTCGGAGCTTAGTGGAGCTTACCTGAACTACCTGGGGCGCGCGCT 300
QY 301 CAACAGAGCTGACTTCAATCCTCTGCTGACCTGGGGCGAGAACTCTGCCAGGGCCACGGT 360
Db 301 CAACAGAGCTGACTTCAATCCTCTGCTGACCTGGGGCGAGAACTCTGCCAGGGCCACGGT 360
QY 361 CAACCTTGAAGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
Db 361 CAACCTTGAAGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
QY 421 GTACAGTCACTTCTGTTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Db 421 GTACAGTCACTTCTGTTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
QY 481 CCGAGCTAGCTGGCGCACTTCTGTATCCAGCTCCAGGCGCTGCTGGCGAGCATTTGCGAG 540
Db 481 CCGAGCTAGCTGGCGCACTTCTGTATCCAGCTCCAGGCGCTGCTGGCGAGCATTTGCGAG 540
```

QY 541 TGTCTGCGGACGCTTGGCTTACCACTGCCCCAGCCCTCTGCCAGGACTGAGCCAGCCTG 600  
DB 541 TGTCTGCGGACGCTTGGCTTACCACTGCCCCAGCCCTCTGCCAGGACTGAGCCAGCCTG 600  
QY 601 GGCCCTTGGCCCTGCGCCACAGTGAATCTTCCAGAGATGGATGACTTCTGGCTGTGTA 660  
DB 601 GGCCCTTGGCCCTGCGCCACAGTGAATCTTCCAGAGATGGATGACTTCTGGCTGTGTA 660  
QY 661 GGAGCTGCAGACTGGCTATGGCGTTTCCAGCAAGGACTTCAACCGGCTTAAGAAGAGAT 720  
DB 661 GGAGCTGCAGACTGGCTATGGCGTTTCCAGCAAGGACTTCAACCGGCTTAAGAAGAGAT 720  
QY 721 GCAGCTTCCAGAGCTTCACTGACCTTCCAGTGAATGGAGGACATGTTTCTGACCTCTGAC 780  
DB 721 GCAGCTTCCAGAGCTTCACTGACCTTCCAGTGAATGGAGGACATGTTTCTGACCTCTGAC 780  
QY 781 CCTTAACCCCCACACTTCCAGGCGCCAGTCACTGTGTGCTT 819  
DB 781 CCTTAACCCCCACACTTCCAGGCGCCAGTCACTGTGTGCTT 819

## RESULT 3

US-09-016-534-4  
; Sequence 4, Application US/09016534  
; Patent No. 6143874  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; APPLICANT: ELLIOTT, GARY S.  
; APPLICANT: SARMIENTO, ULLA  
; APPLICANT: SENALDI, GIORGIO  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,534  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442B  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 819 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 95..769  
; FEATURE:  
; NAME/KEY: mat peptide  
; LOCATION: 176..769  
; FEATURE:  
; NAME/KEY: sig peptide  
; LOCATION: 95..175  
US-09-016-534-4

Query Match 100.0%; Score 819; DB 3; Length 819;  
Best Local Similarity 100.0%; Pred. No. 8.3e-206;  
Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TATTATTAAGCTTCCGCGAGCGCGGCTCGCCCTCCCACTCGGCAGGCTCTGGGAGA 60  
DB 1 TATTATTAAGCTTCCGCGAGCGCGGCTCGCCCTCCCACTCGGCAGGCTCTGGGAGA 60  
QY 61 GGAGCCGCGCCCGCGCGCCCGCCAGCCATGAGACTCCGAGCAGGGGACTCGT 120  
DB 61 GGAGCCGCGCCCGCGCGCCCGCCAGCCATGAGACTCCGAGCAGGGGACTCGT 120  
QY 121 GGAGATGTTAGCTTGGCTATGACAGGTGTGTGGCACTCCCTGCACTGCAAGCTTTAA 180  
DB 121 GGAGATGTTAGCTTGGCTATGACAGGTGTGTGGCACTCCCTGCACTGCAAGCTTTAA 180  
QY 181 TCGCACAGGAGATCCAGGCGCTGGCCCTCCATCCAGAAAACCTATGACCTCACCCGCTA 240  
DB 181 TCGCACAGGAGATCCAGGCGCTGGCCCTCCATCCAGAAAACCTATGACCTCACCCGCTA 240  
QY 241 CCTGGAGATCAACTCCGAGCTTAGCTGGGACCTACTGAACTTACCTGGGGCCCCCTTT 300  
DB 241 CCTGGAGATCAACTCCGAGCTTAGCTGGGACCTACTGAACTTACCTGGGGCCCCCTTT 300  
QY 301 CAACGAGCTGACTTCAATCCTCTCGACTGGGGGCAAGAACTCTGCCAGGGCCACGGT 360  
DB 301 CAACGAGCTGACTTCAATCCTCTCGACTGGGGGCAAGAACTCTGCCAGGGCCACGGT 360  
QY 361 CAACCTTGGAAAGTGGGGAAGCTCAATGACAGGCTGGGCTGAGCCAGAACTATGAGGC 420  
DB 361 CAACCTTGGAAAGTGGGGAAGCTCAATGACAGGCTGGGCTGAGCCAGAACTATGAGGC 420  
QY 421 GTACAGTCACTCTCTGTGTTACTTGGCTGACCTCAACCGCTCAGGCTGCACAGCTGAAC 480  
DB 421 GTACAGTCACTCTCTGTGTTACTTGGCTGACCTCAACCGCTCAGGCTGCACAGCTGAAC 480  
QY 481 CCGAGCTAGCTGGCCCACTTCTGTACCAAGCTTCCAGGGCTGTGTGGGAGCATTTGCAAG 540  
DB 481 CCGAGCTAGCTGGCCCACTTCTGTACCAAGCTTCCAGGGCTGTGTGGGAGCATTTGCAAG 540  
QY 541 TGTCTGCGGACGCTTGGCTACCACTGCCCCAGCCCTCTGCCAGGACTGAGCCAGCCTG 600  
DB 541 TGTCTGCGGACGCTTGGCTACCACTGCCCCAGCCCTCTGCCAGGACTGAGCCAGCCTG 600  
QY 601 GGCCCTTGGCCCTGCGCCACAGTGAATCTTCCAGAGATGGATGACTTCTGGCTGTGTA 660  
DB 601 GGCCCTTGGCCCTGCGCCACAGTGAATCTTCCAGAGATGGATGACTTCTGGCTGTGTA 660  
QY 661 GGAGCTGCAGACTGGCTATGGCGTTTCCAGCAAGGACTTCAACCGGCTTAAGAAGAGAT 720  
DB 661 GGAGCTGCAGACTGGCTATGGCGTTTCCAGCAAGGACTTCAACCGGCTTAAGAAGAGAT 720  
QY 721 GCAGCTTCCAGAGCTTCACTGACCTTCCAGTGAATGGAGGACATGTTTCTGACCTCTGAC 780  
DB 721 GCAGCTTCCAGAGCTTCACTGACCTTCCAGTGAATGGAGGACATGTTTCTGACCTCTGAC 780  
QY 781 CCTTAACCCCCACACTTCCAGGCGCCAGTCACTGTGTGCTT 819  
DB 781 CCTTAACCCCCACACTTCCAGGCGCCAGTCACTGTGTGCTT 819

RESULT 4  
US-08-792-019B-1  
; Sequence 1, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS

```
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..764
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 171..764
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..170
; US-08-792-019B-1

Query Match      81.7%; Score 669.4; DB 1; Length 797;
Best Local Similarity 92.0%; Pred. No. 1.4e-166;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

Qy 5 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCGCCAGAGCTCTGGGAGAGGAG 64
Db 1 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCGCCAGAGCTCTGGGAGAGGAG 60
Qy 65 CGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 124
Db 61 CGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 119
Qy 125 ATGTTAGCTTGCTATGACGCGTGCTGTGCGACCTCCCTGCGAGTGCAGCTCTTAATCGC 184
Db 120 ATGTTAGCTTGCTATGACGCGTGCTGTGCGACCTCCCTGCGAGTGCAGCTCTTAATCGC 179
Qy 185 ACAGAGATCCAGGCGCTCGCGCGCTCCATCCAGAAACCTATGACCTCACCGGCTACCTG 244
Db 180 ACAGGGGACCCAGGCGCTCGCGCGCTCCATCCAGAAACCTATGACCTCACCGGCTACCTG 239
Qy 245 GAGCATCACTCGGAGCTTAGCTGGAGCTTACCTGAACCTACCTGGGGCCCCCTTTCAAC 304
Db 240 GAGCACCAACTCGGAGCTTGCGTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAAC 299
Qy 305 GAGCCTGACTTCAATCCTCCTCGACTGGGGGCGAGAACTCTGCCAGGGCGACGCTCAAC 364
Db 300 GAGCAGACTTCAACCTCTCCCGCTGGGGGCGAGACTCTGCCAGGGCGACTGTGAC 359
Qy 365 TTGGAAGTGTGGGAAGCCTCAATGACAGGCTCGCGCTGACCCAGAACTATGAGGCGTAC 424
Db 360 TTGGAAGTGTGGGAAGCCTCAATGACAAACTGCGGCTGACCCAGAACTACGAGGCGTAC 419
Qy 425 AGTCACCTCTCTGTGTACTTGTGGTGGCTCAACCGTCAGCTGCCAGCTGAACCTCCGA 484
Db 420 AGCCACCTCTCTGTGTACTTGTGGTGGCTCAACCGTCAGCTGCCAGCTGTGAGCTGCCG 479
Qy 485 CGTAGCTGGCCCACTTCTGTATCCAGGCTCCAGGGGCTGTGGGCGAGCAATTGAGGCTGC 544
Db 480 CGAGCCTTCTGTATCCAGGCTCCAGGGGCTGTGGGCGAGCAATTGCGGGCGTC 539
```

```
Qy 545 ATGGCAGCGCTTGGCTACCCACTGCCAGCTCTCCAGGAGCTGAGCCAGCTGGGCC 604
Db 540 ATGGCAGCTCTGGCTACCCACTGCCAGCTCTCCAGGAGCTGAGCCAGCTGGACT 599
Qy 605 CTGGCGCTGCCCAAGTACTTCTCCAGAAAGTGGATGACTTCTGGCTGTGAAGGAG 664
Db 600 CTGGCGCTGCCCAAGTACTTCTCCAGAAAGTGGATGACTTCTGGCTGTGAAGGAG 659
Qy 665 CTGCAGACCTGGCTATGGCTTACGCAAGGACTTCAACCGCTTAAGAGAGATGCGAG 724
Db 660 CTGCAGACCTGGCTATGGCTTACGCAAGGACTTCAACCGCTTAAGAGAGATGCGAG 719
Qy 725 CCTCAGCAGCTTCACTCAGCTCAGCTTGGAGGACATGTTCTGACTCTGACCTT 783
Db 720 CCTCAGCAGCTTCACTCAGCTCAGCTTGGAGGACATGTTCTGACTCTGACCTT 778
```

```
RESULT 5
US-08-988-819-1
; Sequence 1, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUTROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..764
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 171..764
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..170
; US-08-988-819-1
```

```
Query Match      81.7%; Score 669.4; DB 3; Length 797;
Best Local Similarity 92.0%; Pred. No. 1.4e-166;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

Qy 5 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCGCCAGAGCTCTGGGAGAGGAG 64
```







```

; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED
; OTHER INFORMATION: REGION OF >1KB"
US-08-792-019B-3

Query Match 52.5%; Score 429.6; DB 1; Length 5087;
Best Local Similarity 90.4%; Pred. No. 1.8e-103;
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 276 ACCTGAACTACCTGGGGCCCCCTTTCAACGAGCGCTGACTTCAATCTCTCGACTGGGG 335
DB 3363 AGCTGAACTACCTGGGGCCCCCTTTCAACGAGCGAGACTTCAACCTCCCGCTGGGG 3422
QY 336 CAGNAACTCTGCCAGGCGCCACGCTCAACTTGGAGTGTGGCGAGGCTCAATGACAGGC 395
DB 3423 CAGAGACTCTGCCAGGCGCCACTTTCACCTTGGAGGTGTGGCGAAGGCTCAATGACAAAC 3482
QY 396 TGGCGCTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTGTACTTGGTGGCTCA 455
DB 3483 TGGCGCTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTGTACTTGGTGGCTCA 3542
QY 456 ACCGTCAAGGCTGCCAGAGTGAATCCGAGCTAGCTGGGCCCTTCTGTGTACTTGGTGGCTCA 515
DB 3543 ACCGTCAAGGCTGCCAGAGTGAATCCGAGCTAGCTGGGCCCTTCTGTGTACTTGGTGGCTCA 3602
QY 516 AGGCGCTCTGGGCGAGATTGCGAGTGTGCGAGCGCTTGGCTGCGCTACCTGCCCCAGC 575
DB 3603 AGGCGCTCTGGGCGAGATTGCGAGTGTGCGAGCGCTTGGCTGCGCTACCTGCCCCAGC 3662
QY 576 CTCTGCGAGGACTGAGCCAGCTTGGGCCCTTGGGCCCTTGGGCCCTTGGGCCCTTCCAGA 635
DB 3663 CGTCTGCGAGGACTGAGCCAGCTTGGGCCCTTGGGCCCTTGGGCCCTTCCAGA 3722
QY 636 AGATGGATGACTTCTGGCTGCTGAAGAGCTGCGAGCTGCGAGCTTCCAGTCACTGACCTTGG 755
DB 3723 AGATGGATGACTTCTGGCTGCTGAAGAGCTGCGAGCTGCGAGCTTCCAGTCACTGACCTTGG 3782
QY 696 ACTTCAACCGGCTTAAAGAGAGTGCAGCTCCAGAGCTTCCAGTCACTGACCTTGGCTTGG 755
DB 3783 ACTTCAACCGGCTTAAAGAGAGTGCAGCTCCAGAGCTTCCAGTCACTGACCTTGGCTTGG 3842
QY 756 AGGCACATGTTTCTGACCTCTGACCTT 783
DB 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870

RESULT 9
US-08-988-819-3
; Sequence 3, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
```

```

; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED
; OTHER INFORMATION: REGION OF >1KB"
US-08-988-819-3
```

```

Query Match 52.5%; Score 429.6; DB 3; Length 5087;
Best Local Similarity 90.4%; Pred. No. 1.8e-103;
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 276 ACCTGAACTACCTGGGGCCCCCTTTCAACGAGCGCTGACTTCAATCTCTCGACTGGGG 335
DB 3363 AGCTGAACTACCTGGGGCCCCCTTTCAACGAGCGAGACTTCAACCTCCCGCTGGGG 3422
QY 336 CAGNAACTCTGCCAGGCGCCACGCTCAACTTGGAGTGTGGCGAGGCTCAATGACAGGC 395
DB 3423 CAGAGACTCTGCCAGGCGCCACTTTCACCTTGGAGGTGTGGCGAAGGCTCAATGACAAAC 3482
QY 396 TGGCGCTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTGTACTTGGTGGCTCA 455
DB 3483 TGGCGCTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTGTACTTGGTGGCTCA 3542
QY 456 ACCGTCAAGGCTGCCAGAGTGAATCCGAGCTAGCTGGGCCCTTCTGTGTACTTGGTGGCTCC 515
DB 3543 ACCGTCAAGGCTGCCAGAGTGAATCCGAGCTAGCTGGGCCCTTCTGTGTACTTGGTGGCTCC 3602
QY 516 AGGCGCTCTGGGCGAGATTGCGAGTGTGCGAGCGCTTGGCTGCGCTACCTGCCCCAGC 575
DB 3603 AGGCGCTCTGGGCGAGATTGCGAGTGTGCGAGCGCTTGGCTGCGCTACCTGCCCCAGC 3662
QY 576 CTCTGCGAGGACTGAGCCAGCTTGGGCCCTTGGGCCCTTGGGCCCTTGGGCCCTTCCAGA 635
DB 3663 CGTCTGCGAGGACTGAGCCAGCTTGGGCCCTTGGGCCCTTGGGCCCTTCCAGA 3722
QY 636 AGATGGATGACTTCTGGCTGCTGAAGAGCTGCGAGCTGCGAGCTTCCAGTCACTGACCTTGG 695
DB 3723 AGATGGATGACTTCTGGCTGCTGAAGAGCTGCGAGCTGCGAGCTTCCAGTCACTGACCTTGG 3782
QY 696 ACTTCAACCGGCTTAAAGAGAGTGCAGCTCCAGAGCTTCCAGTCACTGACCTTGGCTTGG 755
DB 3783 ACTTCAACCGGCTTAAAGAGAGTGCAGCTCCAGAGCTTCCAGTCACTGACCTTGGCTTGG 3842
QY 756 AGGCACATGTTTCTGACCTCTGACCTT 783
DB 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870
```

## RESULT 10

US-09-016-534-3  
; Sequence 3, Application US/09016534  
; Patent No. 6143874  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; APPLICANT: ELLIOTT, GARY S.  
; APPLICANT: SARNALDI, ULLA  
; APPLICANT: SARNALDI, GIORGIO  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,534  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442B  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5087 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 137..138  
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED"  
; OTHER INFORMATION: REGION OF >1KB"

## US-09-016-534-3

Query Match 52.5%; Score 429.6; DB 3; Length 5087;  
Best Local Similarity 90.4%; Pred. No. 1.8e-103;  
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 276 ACCTGAACCTACCTGGGGCCCCCTTTCAACGAGCGCTGACCTTCAATCTCTCTCGACTGGGGG 335  
Db 3363 AGCTGAACCTACCTGGGGCCCCCTTTCAACGAGCGAGCTTCAACCCCTCCCGCGCTGGGG 3422  
QY 336 CAGAAACTCTGCCAGGGCCAGCGTCAACTTGAAGTGTGGCGAAGCTCAATGACAGGC 395  
Db 3423 CAGAGACTCTGCCAGGGCCAGCTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAAAC 3482  
QY 396 TCGCGGTGACCCAGAACTATAGGCGGTACAGTCACCTCTCTGTGTACTTCGTGGCGCTCA 455  
Db 3483 TCGCGGTGACCCAGAACTACGAGCGCTTCAACCCCTCCCGCGCTGGGG 3542  
QY 456 ACCGTGAGCTGCCAGCTGACCTCGAGGTAGCTGGCCCACTTCTGTACAGCCTCC 515  
Db 3543 ACCGTGAGCTGCCAGCTGCTGAGCTGCGCGCAGCTGGCCCACTTCTGTACAGCCTCC 3602  
QY 515 AGGGCTGTCTGGGCGAGCATTTGAGGTGTCTATGGCGAGCTTGTGGCTACCCACTGCCCGCAGC 575  
Db 3603 AGGGCTGTCTGGGCGAGCATTTGGGGGTCTATGGCAGCTCTGGGCTACCCACTGCCCGCAGC 3662

QY 576 CTCTGCCAGGAGCTGAGCCAGCTGGGGCCCCCTGGGCCCTGCCACACAGTGACTTCTCTCAGA 635  
Db 3663 CGCTGGCTGGGAGCTGAACCCACTTGGACTCTCTGGCCCTGCCACACAGTGACTTCTCTCAGA 3722  
QY 636 AGATGGATGACTTCTGGCTGCTGAAGAGAGCTGCAGACCTGGCTATGGCCCTTCAGCCAAAG 695  
Db 3723 AGATGGAGCAGCTTCTGGCTGCTGAAGAGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAAG 3782  
QY 696 ACTTCAACCGGCTTAAGAGAGAGATGCAGCTCCAGCAGCTTCACTCACCCCTGCACCTTGG 755  
Db 3783 ACTTCAACCGGCTCAAGAAAGAGATGCAGCTCCAGCAGCTTCACTCACCCCTGCACCTTGG 3842  
QY 756 AGGCACATGTTTCTGACCTCTGACCCCT 783  
Db 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870

## RESULT 11

US-09-106-182-7  
; Sequence 7, Application US/09106182  
; Patent No. 6046035  
; GENERAL INFORMATION:  
; APPLICANT: Shi, Yanggu  
; APPLICANT: Ruben, Steve  
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc  
; STREET: 9410 Key West Ave  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: US  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/106,182  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/051,053  
; FILING DATE: 30-JUN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF385  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 396 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-09-106-182-7

Query Match 6.3%; Score 52; DB 3; Length 396;  
Best Local Similarity 85.3%; Pred. No. 8e-05;  
Matches 59; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 716 AAGATGAGCCCTCCAGCAGCTTCAGTCACCTGACCTGGAGGCACTGGTTTCTGACCT 775  
Db 4 ACGAGGAGCCCTCCAGCAGCTTCAGTCACCTGACCTGGGGGCTCATGGCTTCTAACTT 63  
QY 776 CTGACCCCT 783  
Db 64 CTGACCTT 71

## RESULT 12

US-08-665-259-19/c  
; Sequence 19, Application US/08665259  
; Patent No. 6028173  
; GENERAL INFORMATION:  
; APPLICANT: Landes, Gregory M.  
; APPLICANT: Burn, Timothy C.  
; APPLICANT: Connors, Timothy D.  
; APPLICANT: Dackowski, William R.  
; APPLICANT: Van Raay, Terence J.  
; APPLICANT: Klinger, Katherine W.  
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
; METHODS OF MAKING AND USING SAME  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENZYME CORPORATION  
; STREET: One Mountain Road  
; CITY: Framingham  
; STATE: Massachusetts  
; COUNTRY: United States of America  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,259  
; FILING DATE: 17-JUN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dugan, Deborah A.  
; REGISTRATION NUMBER: 37,315  
; REFERENCE/DOCKET NUMBER: IGS-9.1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (508) 872-8400  
; TELEFAX: (508) 872-5415  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6803 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-665-259-19

Query Match 5.3%; Score 43.2; DB 3; Length 6803;  
Best Local Similarity 49.6%; Pred. No. 0.038;  
Matches 111; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
QY 24 CGCGGCTCGCCCTCCACTCCGCGCAGCCTCTGGGAGAGGAGCGCGCGCGCGCGCGCG 83  
DB 6708 CCCGCCCCCAGCTCCCGCCCCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCG 6649  
QY 84 CCCCAGCCCCCAGCCTCCGAGCAGGGGACTCTGTGGGGGATGTTAGCTTTCCTATGCA 143  
DB 6648 CCGGCAACCCG 6589  
QY 144 CGGTGCTGTGGACCTCCCTGAGTGCAGCTCTTAATCGCACAGGAGATCCAGGCGCTG 203  
DB 6588 CGCGGCTTTCCTCAGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6529  
QY 204 GCCCTCCATCCAGAAACCTATGACCTCACCGCTACTGTGAG 247  
DB 6528 CCCCAGCGAGCTGTGCGCCCTGTGCCACCCCGGAGAGCGCGCG 6485

## RESULT 13

US-08-762-500-19/c  
; Sequence 19, Application US/08762500  
; Patent No. 6030806  
; GENERAL INFORMATION:

APPLICANT: Landes, Gregory M.  
APPLICANT: Burn, Timothy C.  
APPLICANT: Connors, Timothy D.  
APPLICANT: Dackowski, William R.  
APPLICANT: Van Raay, Terence J.  
APPLICANT: Klinger, Katherine W.  
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
METHODS OF MAKING AND USING SAME  
NUMBER OF SEQUENCES: 83  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENZYME CORPORATION  
STREET: One Mountain Road  
CITY: Framingham  
STATE: Massachusetts  
COUNTRY: United States of America  
ZIP: 01701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/762,500  
FILING DATE: 09-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/665,259  
FILING DATE: 17-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10469  
FILING DATE: 17-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.  
REGISTRATION NUMBER: 37,315  
REFERENCE/DOCKET NUMBER: IGS-9.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508) 872-8400  
TELEFAX: (508) 872-5415  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6803 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-762-500-19

Query Match 5.3%; Score 43.2; DB 3; Length 6803;  
Best Local Similarity 49.6%; Pred. No. 0.038;  
Matches 111; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
QY 24 CGCGGCTCGCCCTCCACTCCGCGCAGCCTCTGGGAGAGGAGCGCGCGCGCGCGCG 83  
DB 6708 CCCGCCCCCAGCTCCCGCCCCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCG 6649  
QY 84 CCCCAGCCCCCAGCCTCCGAGCAGGGGACTCTGTGGGGGATGTTAGCTTTCCTATGCA 143  
DB 6648 CCGGCAACCCG 6589  
QY 144 CGGTGCTGTGGACCTCCCTGAGTGCAGCTCTTAATCGCACAGGAGATCCAGGCGCTG 203  
DB 6588 CGCGGCTTTCCTCAGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6529  
QY 204 GCCCTCCATCCAGAAACCTATGACCTCACCGCTACTGTGAG 247  
DB 6528 CCCCAGCGAGCTGTGCGCCCTGTGCCACCCCGGAGAGCGCGCG 6485

## RESULT 14

US-08-665-037-1  
; Sequence 1, Application US/08665037  
; Patent No. 5895813  
; GENERAL INFORMATION:

; APPLICANT: Seedorf, Klaus  
; APPLICANT: Ullrich, Axel  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT  
; TITLE OF INVENTION: OF TKA-1 RELATED  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,037  
; FILING DATE: June 13, 1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/005,167  
; FILING DATE: October 13, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 220/156  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1642 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 97...1446  
US-08-665-037-1

Query Match 5.2%; Score 42.4; DB 2; Length 1642;  
Best Local Similarity 52.9%; Pred. No. 0.04;  
Matches 91; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 512 CTCACGGGCTGCTGGGCGACGATTGCGAGGTGTATGGCGAGCTTGGCTACCCACTGCC 571  
Db 379 CTCGCGCGGCGGCGAGCTGACCTGTACCGAGGAGTGGCCCGAGGAGGCTCCACCGCC 438  
Qy 572 CAGCCTCTCCAGGAGTACGACGAGCTGGGCGCCCTGGCCCTGCCACAGTACTTCTTC 631  
Db 439 CACGACCCCTGGAGCGGAGCGGAGCTGGGCGACACACCGGCGAGCTCCGAGCT 498  
Qy 632 CAGAAGATGATGACTTCTGGCTGTGAAGGAGCTGCGAGCTTGGCTATGGC 683  
Db 499 GCGAAGAGGATGTAGTGGGCGCCCTGAGGAGCTGGGCGCTCGGCTCTGCC 550

RESULT 15  
US-08-666-067-1  
; Sequence 1, Application US/08666067  
; Patent No. 5922842  
; GENERAL INFORMATION:  
; APPLICANT: Seedorf, Klaus  
; APPLICANT: Ullrich, Axel  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT  
; TITLE OF INVENTION: OF TKA-1 RELATED  
; TITLE OF INVENTION: DISORDERS

; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/666,067  
; FILING DATE: June 13, 1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/005,421  
; FILING DATE: October 13, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 220/157  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1642 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 97...1446  
US-08-666-067-1

Query Match 5.2%; Score 42.4; DB 2; Length 1642;  
Best Local Similarity 52.9%; Pred. No. 0.04;  
Matches 91; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 512 CTCACGGGCTGCTGGGCGACGATTGCGAGGTGTATGGCGAGCTTGGCTACCCACTGCC 571  
Db 379 CTCGCGCGGCGGCGAGCTGACCTGTACCGAGGAGTGGCCCGAGGAGGCTCCACCGCC 438  
Qy 572 CAGCCTCTGCCAGGAGTACGACGAGCTGGGCGCCCTGGCCCTGCCACAGTACTTCTTC 631  
Db 439 CACGACCCCTGGAGCGGAGCGGAGCTGGGCGACACACCGGCGAGCTCCGAGCT 498  
Qy 632 CAGAAGATGATGACTTCTGGCTGTGAAGGAGCTGCGAGCTTGGCTATGGC 683  
Db 499 GCGAAGAGGATGTAGTGGGCGCCCTGAGGAGCTGGGCGCTCGGCTCTGCC 550

Search completed: February 1, 2003, 08:50:20  
Job time : 46.3918 secs

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:22 ; Search time 14.5 Seconds  
(without alignments)  
1491.741 Million cell updates/sec

Title: US-09-931-704-5  
Perfect score: 1224  
Sequence: 1 MDLRAGDSWGLACLCTVLW.....KKKMQPPAASVTLHLEAHGF 225

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118.5	9.7	203	2 JC4645	cardiotrophin-1 -
2	110.5	9.0	203	2 I49153	cardiotrophin-1 -
3	93.5	7.6	640	2 T34916	transferase - Stre
4	92.5	7.6	200	2 AD3633	hypothetical prote
5	92	7.5	332	2 G86182	hypothetical prote
6	91.5	7.5	201	2 G02312	cardiotrophin-1 -
7	90	7.4	195	2 JH0680	ciliary neurotroph
8	90	7.4	733	2 F82965	hypothetical prote
9	89	7.3	733	2 D83588	conserved hypotet
10	86	7.0	1313	2 T38943	probable guanine n
11	85.5	7.0	389	2 F72511	probable cystathio
12	85.5	7.0	560	2 C38604	poly(3-hydroxyalka
13	85	6.9	520	2 B75332	extracellular solu
14	84	6.9	741	2 A83271	hypothetical prote
15	83	6.8	272	2 T35231	hypothetical prote
16	83	6.8	955	2 T00247	zinc finger protei
17	83	6.8	1561	2 T00248	zinc finger protei
18	80.5	6.6	1006	2 T42762	probable alpha-man
19	80	6.5	382	2 AB2196	two-component hybr
20	80	6.5	1179	2 T04584	TWV resistance pro
21	79.5	6.5	224	2 S32947	hupK protein - Rho
22	79.5	6.5	316	2 T34553	hypothetical prote
23	79.5	6.5	512	2 S21171	activin receptor S
24	79.5	6.5	542	2 A82965	hypothetical prote
25	79.5	6.5	644	2 JC5119	hypothetical prote
26	79	6.5	1132	2 A35098	anti-mullerian hor
27	78.5	6.4	338	2 T45394	MHC class III hist
28	78.5	6.4	400	2 AF0873	hypothetical prote
29	78.5	6.4	522	2 D87123	probable monooxyge
					conserved hypotet

30 78.5 6.4 531 2 B83422 probable serine/th  
31 78.5 6.4 571 2 H70623 probable potassium  
32 78.5 6.4 1844 2 S01956 hypothetical prote  
33 78 6.4 232 2 AG2157 hypothetical prote  
34 78 6.4 378 2 D83381 hypothetical prote  
35 78 6.4 427 1 RRYC65 RNA-directed DNA p  
36 78 6.4 515 2 T37982 hypothetical prote  
37 78 6.4 917 2 S40178 isoleucine-tRNA li  
38 77.5 6.3 572 2 D89891 ile-tRNA synthetas  
39 77.5 6.3 1182 2 I48378 probable methionyl  
40 77.5 6.3 322 2 T23891 hairless protein -  
41 77 6.3 327 2 C97663 hypothetical prote  
42 77 6.3 327 2 AF2887 conserved hypotet  
43 77 6.3 411 2 F75439 probable cell wall  
44 77 6.3 1031 2 F83561 probable DNA polym

## ALIGNMENTS

## RESULT 1

JC4645  
cardiotrophin-1 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Jun-2000  
C:Accession: JC4645  
R:Ishikawa, M.; Saito, Y.; Miyamoto, Y.; Kuwahara, K.; Ogawa, E.; Nakagawa, O.; Harada  
Biochem. Biophys. Res. Commun. 219, 377-381, 1996  
A:Title: cDNA cloning of rat cardiotrophin-1 (CT-1): Augmented expression of CT-1 gen  
A:Reference number: JC4645; MUID:96193659; PMID:9604995  
A:Accession: JC4645  
A:Molecule type: mRNA  
A:Residues: 1-203 <ISH>  
A:Cross-references: DDBJ:D78591; NID:G1256926; PIDN:BA11427.1; PID:G1256927  
C:Genetics:  
A:Gene: CT-1  
C:Keywords: cardiac muscle; cytokine; heart

Query Match 9.7%; Score 118.5; DB 2; Length 203;  
Best Local Similarity 27.5%; Pred. No. 0.0015;  
Matches 49; Conservative 26; Mismatches 84; Indels 19; Gaps 6;  
Qy 40 IQKTYDLRYLEHQRLSLAGTYLVNLYGPPENEDPBNRL---GAETLPRAVTVNLEWRS 96  
Db 27 IRQTHNLARLLTKYADQLLEEVVQQQGFGLGFPFRLPLAGUSGPAPSHAGLPV--- 83  
Qy 97 LNDRLRLTQNYEAVSHLLCYLRGLNRQAA-----TAELEERSLAHFCTSLQGLLGSTAGYM 151  
Db 84 -SERLR--QDAALASALPALLDAVERRQALNPAPRLRLSLSDAARQVRALGAATVTL 140  
Qy 152 ATLGV---PLPQPLPGTEPAWAGPAHSDFLOKMDDFWLLKELQTLWLRSAKDFNRL 205  
Db 141 AALGAARAGPVPEPV-ATSAFTNSAAGVFSKVLGLHVGCLYGEWVSRTEGDLGQL 197

## RESULT 2

I49153

cardiotrophin-1 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999

C:Accession: I49153

R:Pennica, D.; King, K.L.; Shaw, K.J.; Luis, E.; Rullamas, J.; Luoh, S.; Darbonne, W.C.

Proc. Natl. Acad. Sci. U.S.A. 92, 1142-1146, 1995

A:Title: Expression cloning of cardiotrophin 1, a cytokine that induces cardiac myocyte

A:Reference number: I49153; MUID:95166785; PMID:7862649

A:Accession: I49153

A>Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-203 <RES>

A:Cross-references: EMBL:U18366; NID:G710331; PIDN:AAC52173.1; PID:G710332

C:Genetics:  
A:Gene: ctfl



```
Db 267 DDPDESADKEDAEVQKL-----SEKDL--LKRHIE 298

RESULT 6
G02312
cardiotrophin-1 - human
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C:Accession: G02312
R:Wood, W.I.
submitted to the EMBL Data Library, December 1995
A:Reference number: H01035
A:Accession: G02312
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-201 <WOO>
A:Cross-references: EMBL:U40303; NID:gl151149; PID:gl151150
C:Genetics:
A:Gene: GDB:CTF1; CT-1
A:Cross-references: GDB:567078
A:Map position: lp22-lp22

Query Match 7.5%; Score 91.5; DB 2; Length 201;
Best Local Similarity 25.7%; Pred. No. 0.47;
Matches 45; Conservative 23; Mismatches 92; Indels 15; Gaps 5;

QY 40 IQKTYDLTRYLEHQLSLAGTYLNYLGGPPNEDFPNPRIGATLPRTATVNVLEWVRSND 99
Db 27 IQTHSLAHLTKYABQLQYVQLQDPPGLPSPRPPLPVAGL-SAPAPSHAGLPVHE 85

QY 100 RLRLTONYEAYSHLLCYLRLNRQAA-----TAELESLAHFCTSLQGLLGSTAGVWATL 154
Db 86 RLRL--DAAALAPLLLDVACRQELNPRAPRLRLLEDAARQARALGAAVEALLAAL 143

QY 155 GY-----PLPQPLPGTEPAWAPGAHSDFLQKMDDFWLLKELOTWLRSAKDFNRL 205
Db 144 GAANRGPRAEPPAATASA---ASATGVFPAKVLGLRVCGLYREWLSTRTGDLGQL 195

RESULT 7
JH0680
ciliary neurotrophic factor - chicken
N:Alternate names: growth-promoting activity protein
C:Species: Gallus gallus (Chicken)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
C:Accession: JH0680; PQ0057
R:Leung, D.W.; Parent, A.S.; Cachianes, G.; Esch, F.; Coulombe, J.N.; Nikolics, K.; Ecker,
Neuron 6, 1045-1053, 1992
A:Title: Cloning, expression during development, and evidence for release of a trophic f
A:Reference number: JH0680; MUID:92304573; PMID:1610564
A:Accession: JH0680
A:Molecule type: mRNA
A:Residues: 1-195 <LEU>
A:Cross-references: GB:M0827; NID:g211822; PIDN:AAA48784.1; PID:g211823
A:Experimental source: eye
R:Eckenstein, F.P.; Esch, F.; Holbert, T.; Blacher, R.W.; Nishi, R.
Neuron 4, 623-631, 1990
A:Title: Purification and characterization of a trophic factor for embryonic peripheral
A:Reference number: PQ0057; MUID:90211978; PMID:2322465
A:Accession: PQ0057
A:Molecule type: protein
A:Residues: 155-166, 'X', 168-175 <ECK>
A:Experimental source: sciatic nerves
C:Comment: This is a neurotrophic protein.
C:Superfamily: ciliary neurotrophic factor
C:Keywords: growth factor

Query Match 7.4%; Score 90; DB 2; Length 195;
Best Local Similarity 27.7%; Pred. No. 0.63;
Matches 52; Conservative 21; Mismatches 83; Indels 32; Gaps 10;

QY 46 LTRYLEHQLSLAGTYLNYLGGPPNEDFPNPRIGATLPRTATVNVLEWVRSNDRLRLTQ 105

Db 23 LARKWRSVDTLDDLIYVERQG-----LDASISVAADVGVPTAAV--ERWAEQTGTQRLLD 75

QY 106 N---YEAYSHLLCYLRLNRQA---ATAELRRSLA-----HFCTSLQGLLGSTAGVMA 152
Db 76 NLAAAYRAFRTLLAQMEEQRELLGDTDAELGPALAAMLLQVSAFVYHLEELL-----ELE 130

QY 153 TLGVPLPQPLPGTEPAWAPGAHSDFLQKMDDFWLLKELOTWLRSAKDFNRLKKMQP 212
Db 131 SRGAPAE---GSEPP-AP-PRUSLFQKLRGLRVLELAQWAVRSVRDLRLQLSK--HGP 183

QY 213 AASVTLHL 220
Db 184 GSCAALGL 191

RESULT 8
F82965
hypothetical protein PA5441 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F82965
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F82965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-733 <STO>
A:Cross-references: GB:AE004957; GB:AE004091; NID:g9951770; PIDN:AAG08826.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA5441

Query Match 7.4%; Score 90; DB 2; Length 733;
Best Local Similarity 23.9%; Pred. No. 3.2;
Matches 48; Conservative 24; Mismatches 81; Indels 48; Gaps 8;

QY 23 PAVPALNRTGDPGPGSPQTKTYDLTRYLEHQLSLAGTYLNYLGGPPNEDFPNPRIGAE 82
Db 102 PALPAASEAETPPAPAPSPPLAEIAR-----QMGAEL 132

QY 83 TLPRATVNVLEWVRSNDRLRLTONYEAYSHLLCYLRLNRQAATAELR-RSLAHFCTSLQ 141
Db 133 ALPEXIAGAEFFEGSGRCRSNDQSA---LAFRLQV-RDAGLGEAETKALAN---SRL 184

QY 142 GLLGSIAGVWATLGYPLPQPLPGTEPAWAPGAHSDFLQKMDDFWLLKELOTWLRSAK 201
Db 185 DLLGACGWEQBELGGLVLAQ---GVES--AAGKAFATYLEAANAFYSGR-----FDEAEQG 234

QY 202 FNRLKKMKQPPAASVTLHLEA 222
Db 235 FKALQDVSQPMKLKETALYLQA 255

RESULT 9
D83588
conserved hypothetical protein PA0454 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83588
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pat
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: D83588
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-733 <STO>
```

A;Cross-references: GB:AE004483; GB:AE004091; NID:G9946313; PIDN:AG03843.1; GSPDB:GN001  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA0454  
C;Superfamily: hypothetical protein H11680

Query Match 7.3%; Score 89; DB 2; Length 733;  
Best Local Similarity 25.8%; Pred. No. 4;  
Matches 57; Conservative 26; Mismatches 64; Indels 74; Gaps 13;

Qy 28 LNRTGDPGPGSIQTYDITRYLE-----HQLSLAGTYLNYLGGPPNEPD--FNPP 77  
Db 230 LNLRLGHRGPGPKV-----SRVLYFIADQVHERASSHPYNNRLAEAFHSDVLFRCQ 283  
Qy 78 RL-----GAETLPRTAVNLVWRSNDRLRTQNYEAYSHLLCYLRGLNRQAATAELELR 131  
Db 284 RLNLQGGKACQALARA-----IRLQRPD-YA-----DRELALEDLQA 320  
Qy 132 SLAHP-----CTSLQGL-----LGSAGVWATLGYLPQPLPGTEPAWAPGPAHSDFLQKMD 184  
Db 321 SLEHLRQQSNPAWKGLRLSGLAALNTLDRKL-----AGASNPDIADEQD 369  
Qy 185 FWLKEQLTWLWESAKD-FNRLKKMKOPPAASVTLLHLEAHG 224  
Db 369 SALLDRSP-----RSLKDAFERLRQUTP-----TSLLFRHG 400

RESULT 10  
T38943  
probable guanine nucleotide binding protein - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C;Accession: T38943  
R;Skellton, J.; Church, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, May 1997  
A;Reference number: Z21819  
A;Accession: T38943  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1313 <SKE>  
A;Cross-references: EMBL:Z95396; PIDN:CA08769.1; GSPDB:GN00066; SPDB:SPAC57A7.11  
A;Experimental source: strain 972h-; cosmid c57A7  
C;Genetics:  
A;Gene: SPDB:SPAC57A7.11  
A;Map position: 1

Query Match 7.0%; Score 86; DB 2; Length 1313;  
Best Local Similarity 22.2%; Pred. No. 16;  
Matches 58; Conservative 31; Mismatches 84; Indels 88; Gaps 12;

Qy 15 LCTVLMHLPV--PALNR-----TGDPGPGPSI-----QKTYDITRYLEHQLRS 56  
Db 811 LAFILLOLPAHKAASLSDKDTNSVDPKHPFVPSVSENKILNRSFSLTRSLKGLALS 870  
Qy 57 LAG-----TYLNL-----GPP-FNEPDNPPRLGAETLPRTAVNL 91  
Db 871 LAGDRASELLSNGENKPAESNLHLTSAKVPGPAPNELEYQ-----SELDMLPTSYLF 926  
Qy 92 EVWRSNDRRLRTQNYEAYSHLLCYLRGL-----NRQAATAELRSLAHFCTS-----LOG 142  
Db 927 DWSRKYFTEPQWRNEDDEPGSICYNQRLWRNRNEKLIYRTPLAEYSTNGRWNQQLMT 986  
Qy 143 LLGSIAGVWATLGYLPQPLPGTEPAWAPGPAHSDFLQKMDFFLLKE---LQTLWLRSA 199  
Db 987 FNNTIA-----PRKLMFHQPEDQLITLGDKDIIVQWDRRN 1022  
Qy 200 KDFNRLKKMKOPPAASVTLLH 220  
Db 1024 RCLNSFK-----TSASATNV 1038

RESULT 11  
F72511  
probable cystathionine gamma-synthase APE2068 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C;Accession: F72511  
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yanazaki, S.; Haikawa, Y.; Jin-no, K.; Tawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.  
DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix  
A;Reference number: A72450; MUID:99310339; PMID:10382966  
A;Accession: F72511  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-389 <RAW>  
A;Cross-references: DDBJ:AP000063; NID:G5105654; PIDN:BAA81078.1; PID:G5105766  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE2068  
C;Superfamily: O-succinylhomoserine (thiol)-lyase

Query Match 7.0%; Score 85.5; DB 2; Length 389;  
Best Local Similarity 22.9%; Pred. No. 3.9;  
Matches 47; Conservative 20; Mismatches 77; Indels 61; Gaps 8;

Qy 43 TYDITRYLEHQLRSAGTYLNYLGGPPNEPDNPPRLGAETLPRTAVNLVWRS----- 97  
Db 107 TYGSTRSLEMLSSITGIEVRLAGPPWED-----LLDLVCWADLIIVES 150  
Qy 98 --NDRRLRTQNYEAYSHLLCYLRGLNRQAATAELR-----RSL-----AHFCTS 139  
Db 151 MANPTLRVPP-----LSGIYRAGSCGVVVDNTFATPLAYRLEGAHY--\$ 197  
Qy 140 LQGLLGSAGVWATLGYLPQPL--PGTEPAWAPGPAHSDFLQKMDFFLLKELQTLWRS 198  
Db 198 LESLTKYTAGHNDVVGSLSGRVEDLEPLNMNRKILGTIMQPIDAY-----LAWRG 249  
Qy 199 AKDFNRLKKMKOPPAASVTLLHLEAH 223  
Db 250 MKTLKARFEAQSRAAVAEWLESH 274

RESULT 12  
C38604  
poly(3-hydroxyalkanoate) polymerase (EC 2.7.7.-) 3 - Pseudomonas oleovorans  
C;Species: Pseudomonas oleovorans  
C;Date: 23-Aug-1991 #sequence\_revision 23-Aug-1991 #text\_change 08-Sep-2000  
C;Accession: C38604  
R;Huisman, G.W.; Wonink, E.; Meima, R.; Kazemier, B.; Terpstra, P.; Witholt, B.  
J. Biol. Chem. 266, 2191-2198, 1991  
A;Title: Metabolism of poly(3-hydroxyalkanoates) (PHAs) by Pseudomonas oleovorans. I. Characterization of the poly(3-hydroxyalkanoate) synthase gene  
A;Reference number: A38604; MUID:91115830; PMID:1989978  
A;Accession: C38604  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-560 <HUI>  
A;Cross-references: GB:M58445; NID:G151441; PIDN:AAA25934.1; PID:G151444  
C;Superfamily: poly(3-hydroxyalkanoic acid) synthase phbc  
C;Keywords: nucleotidyltransferase

Query Match 7.0%; Score 85.5; DB 2; Length 560;  
Best Local Similarity 30.2%; Pred. No. 6;  
Matches 38; Conservative 16; Mismatches 43; Indels 29; Gaps 7;

Qy 77 PRIGATLPRTAVNLVWRSNDRRLRTQNYEAYSHLLCYLRGLNRQAATAELRSL--A 134  
Db 5 PAKGTPTLPATSMNVQ-----NAILGLR-----DLISLRNVSROS-----LRHPLHTA 50  
Qy 135 HFCTSLQGLSAGVWATLGYLPQPLPG-----TEPAWAPGPAHSDFLQKMDFFLLKE 190  
Db 51 HHLLALGGQLGRV-----ILGDTPLQNPDRPSDFTNSQNPYRGLQA-----YLAQW 101  
Qy 191 LQTLW 196  
Db 102 KQTLW 107



	Query Match	6.9%	Score 84;	DB 2;	Length 741;
	Best Local Similarity	26.2%;	Pred. No. 12;		
	Matches 64;	Conservative 13;	Mismatches 81;	Indels 86;	Gaps 13;
5	AGDSW----	GMLACLCITVLWHLPV	ALNRRCTDCPGPSIQKTY	LDLTRLEHQLRS	LAGT 60
447	AAAWSAVALGMLGALC--	WLAPA	-----GLP-----		
61	YLVYLGPPFNPDPNPPLGAET	LP	PRATVN--LEVWRS	INDRLRLTQNYEAYSHLLC	YLRLG 119









RT "Cloning, expression during development, and evidence for release of  
RT a trophic factor for ciliary ganglion neurons.";  
RL Neuron 8:1045-1053(1992).

CC -1- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL  
CC TYPES. SEEMS TO PREVENT THE DEGENERATION OF MOTOR AXONS AFTER  
CC AXOTOMY.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM.

CC -1- SIMILARITY: BELONGS TO THE CNTF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; M80827; AAA48784.1; -;  
CC PIR; JH0680; JH0680.

CC HSPF; P26441; ICNT.

CC InterPro; IPR000151; Ctl\_neuro\_factor.

CC Pfam; PF01110; CNTF; 1.

CC ProDom; PD011041; Ctl\_neuro\_factor; 1.

CC Growth factor; Neurone.

CC SEQUENCE 195 AA; 21330 MW; FEA076949DB34AC5 CRC64;

Query Match 7.4%; Score 90; DB 1; Length 195;

Best Local Similarity 27.7%; Pred. No. 0.23;

Matches 52; Conservative 21; Mismatches 83; Indels 32; Gaps 10;

QY 46 LTRYLEQLRLSLAGTYLVYLPFPFNEPDPNPPPLGAEITLPRATVNLVWRSNDRLRLTQ 105

DB 23 LARKMRSDVTDLDIYVERQG-----LDASISVAADVGVPTAAV--ERMAEQGTORLLD 75

QY 106 N---YEAYSHLLCYLRLGNLROA---ATAELRRSLA-----HFCTSLQGLGSIAGVMA 152

DB 76 NLAAVPAFTLLAQMLEEQRELLGDTDAELGPALAAWLLQSAFVYHLELL-----ELE 130

QY 153 TLGYPLPQPLPGTEPAWAPGPAHSDFLQKMDDFWLKELQTLWLRSAKDNRLKKMQPP 212

DB 131 SRGAPAE---GSEPP-AP-PRLSLFEQKLRGLRLVRLAQWAVRSVRDLRLQLSK--HGP 183

QY 213 AASVTLHL 220

DB 184 GSGAALGL 191

RESULT 6

M2B2 HUMAN

ID M2B2 HUMAN STANDARD; PRT; 1009 AA.

AC Q9Y2E5;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE Epididymis-specific alpha-mannosidase precursor (EC 3.2.1.24)

DE (Mannosidase alpha class 2B member 2).

GN MAN2B2 OR KIA00935.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]\_TaxID=9606;

RP SEQUENCE FROM N.A.

RA Stone N.E., Schmutz J.J., Cox D.R., Myers R.M.;

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 132-1009 FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=99246063; PubMed=10231032;

RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,

RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XIII.

RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:63-70(1999).

RN [3]

RP RECONSTRUCTION FROM GENOMIC SEQUENCE.

RA Bairoch A.;

RL Unpublished observations (NOV-2001).  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-  
CC mannose residues in alpha-D-mannosides.

CC -1- SUBCELLULAR LOCATION: Secreted (Potential).

CC -1- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; AC004480; -; NOT ANNOTATED CDS.

CC EMBL; AB023152; BAA76779.1; ALT SEQ.

CC Hydrolase; Glycosidase; Signal; Glycoprotein.

FT SIGNAL 1 23

FT CHAIN 24 1009

FT CARBOHYD 226 226

FT CARBOHYD 249 249

FT CARBOHYD 294 294

FT CARBOHYD 336 336

FT CARBOHYD 516 516

FT CARBOHYD 608 608

FT CARBOHYD 670 670

FT CARBOHYD 675 675

FT CARBOHYD 748 748

FT CARBOHYD 808 808

FT CARBOHYD 812 812

FT CARBOHYD 890 890

FT SEQUENCE 1009 AA; 113987 MW; F3DB81DD061352E6 CRC64;

Query Match 7.2%; Score 88; DB 1; Length 1009;

Best Local Similarity 23.3%; Pred. No. 2.8;

Matches 45; Conservative 21; Mismatches 65; Indels 62; Gaps 8;

QY 50 LEHQ-----LRSLAGTYLVYLPFPFNEPDPNPPPLGAEITLPRATVNLVWRSNDRLRLTQ 105

DB 842 LQHRPVVLFGDLGAPKLPQGPQOEAVTLPNHLQIL-----SIPGWRYSNHTHSQ 896

QY 106 NYEAYSHLLCYLRLGNLROAATAELRRSLAHF-----CTSLQGLGSIAG 149

DB 897 N-----LRKHGRGEAQADLRVLLRLVHLYEVGEDPVLSPQPTVNLQVLAQGS 946

QY 150 VMATLGYPLPQPLPGTEPAWAPGPAHSDFLQKMDDFWLKELQTLWLRSAKDNFNR--LKK 207

DB 947 VVAV-----EERSLGT-----WDLMLHRWSWRTGPGHRGDTTS 982

QY 208 KMOPPAAS-VTLH 219

DB 983 PSRPPGGPIITVH 995

RESULT 7

NXF1 HUMAN

ID NXF1 HUMAN STANDARD; PRT; 619 AA.

AC Q9UBU9; Q9UQL2; Q99799;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE Nuclear RNA export factor 1 (Tip associating protein) (Tip-associated

DE protein) (mRNA export factor TAP).

GN NXF1 OR TAP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
RC TISSUE=Cervical carcinoma;  
RX MEDLINE=99219873; PubMed=10202158;  
RA Braun I.C., Rohrbach E., Schmitt C., Izaurralde E.;  
RT "TAP binds to the constitutive transport element (CTE) through a novel  
RNA-binding motif that is sufficient to promote CTE-dependent RNA  
export from the nucleus.";  
RL EMBO J. 18:1953-1965(1999).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99257272; PubMed=10323864;  
RA Kang Y., Cullen B.R.;  
RT "The human Tap protein is a nuclear mRNA export factor that contains  
novel RNA-binding and nucleocytoplasmic transport sequences.";  
RL Genes Dev. 13:1126-1139(1999).  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93384298; PubMed=10454577;  
RA Bear J., Tan W., Zolotukhin A.S., Tabernero C., Hudson E.A.,  
RA Felber B.K.;  
RT "Identification of novel import and export signals of human TAP, the  
protein that binds to the constitutive transport element of the type  
D retrovirus mRNA.";  
RL Mol. Cell. Biol. 19:6306-6317(1999).  
[4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBSJ databases.  
[5]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,  
RA Yamada K., Fujii Y., Oraki K., Hirao M., Ohmori Y., Oka T., Suzuki Y.,  
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,  
RA Isegai T., Sugano S.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBSJ databases.  
[6]  
RP SEQUENCE OF 61-619 FROM N.A.  
RC TISSUE=Lymphocytes;  
RX MEDLINE=97318898; PubMed=9175835;  
RA Yoon D.-W., Lee H., Seol W., DeMaria M., Rosenzweig M., Jung J.U.;  
RT "Tap: a novel cellular protein that interacts with tip of herpesvirus  
saimiti and induces lymphocyte aggregation.";  
RL Immunity 6:571-582(1997).  
[7]  
RP FUNCTION.  
RX PubMed=9660949;  
RA Grueter P., Tabernero C., von Kobbe C., Schmitt C., Saavedra C.,  
RA Bachi A., Wilm M., Felber B.K., Izaurralde E.;  
RT "TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export  
from the nucleus.";  
RL Mol. Cell 1:649-659(1998).  
[8]  
RP CHARACTERIZATION.  
RX MEDLINE=21282872; PubMed=11259411;  
RA Braun I.C., Herold A., Rode M., Conti E., Izaurralde E.;  
RT "Overexpression of TAP/p15 heterodimers bypasses nuclear retention and  
stimulates nuclear mRNA export.";  
RL J. Biol. Chem. 276:20536-20543(2001).  
[9]  
RP CHARACTERIZATION.  
RX PubMed=10668806;  
RA Bachi A., Braun I.C., Rodrigues J.P., Pante N., Ribbeck K.,  
RA von Kobbe C., Kutay U., Wilm M., Gorlich D., Carmo-Fonseca M.,  
RA Izaurralde E.;  
RT "The C-terminal domain of TAP interacts with the nuclear pore complex  
and promotes export of specific CTE-bearing RNA substrates.";  
RL RNA 6:136-158(2000).

[10]  
RN MUTAGENESIS.  
RP MEDLINE=21151125; PubMed=11256625;  
RA Suyama M., Doerks T., Braun I.C., Sattler M., Izaurralde E., Bork P.;  
RT "Prediction of structural domains of TAP reveals details of its  
interaction with p15 and nucleoporins.";  
RL EMBO Rep. 1:53-58(2000).  
[11]  
RN X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 102-372.  
RX MEDLINE=20514125; PubMed=11060011;  
RA Likier E., Fernandez E., Izaurralde E., Conti E.;  
RT "The structure of the mRNA export factor TAP reveals a cis arrangement  
of a non-canonical RNP domain and an LRR domain.";  
RL EMBO J. 19:5587-5598(2000).  
[12]  
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH NXT1, AND X-RAY  
RP CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF COMPLEX WITH NXT1-FG-REPEAT.  
RX PubMed=11583626;  
RA Fribourg S., Braun I.C., Izaurralde E., Conti E.;  
RT "Structural basis for the recognition of a nucleoporin FG repeat by  
the NTF2-like domain of the TAP/p15 mRNA nuclear export factor.";  
RL Mol. Cell 8:645-656(2001).  
[13]  
RN STRUCTURE BY NMR OF 551-619, AND MUTAGENESIS OF PHE-617.  
RX MEDLINE=21912422; PubMed=11875519;  
RA Grant R.P., Hurt E., Neuhaus D., Stewart M.;  
RT "Structure of the C-terminal FG-nucleoporin binding domain of  
Tap/NXF1.";  
RL Nat. Struct. Biol. 9:247-251(2002).  
CC -!- FUNCTION: Involved in the nuclear export of mRNA species bearing  
retroviral constitutive transport elements (CTE) and in the export  
of mRNA from the nucleus to the cytoplasm.  
CC -!- SUBUNIT: Interacts with NXT1, NXT2, E1B-AP5, RAE1, the REF  
proteins and with several nucleoporins  
CC -!- SUBCELLULAR LOCATION: Nuclear; localized predominantly in the  
nucleoplasm and at both the nucleoplasmic and cytoplasmic faces of  
the nuclear pore complex. Shuttles between the nucleus and the  
cytoplasm.  
CC -!- TISSUE SPECIFICITY: Expressed ubiquitously.  
CC -!- DOMAIN: The minimal CTE binding domain consists of an RNP-  
type RNA binding domain (RBD) and leucine-rich repeats.  
CC -!- DOMAIN: The nucleoporin binding domain consists of a NTF2-like  
domain and a UBA-like domain. The NTF2 domain heterodimerizes with  
NXT1 AND NXT2. The formation of NXF1/NXT1 heterodimers is required  
for NXF1-mediated nuclear mRNA export. The UBA-like domain  
mediates direct interactions with nucleoporin-FG-repeats and is  
necessary and sufficient for localization of NXF1 to the nuclear  
rim. The conserved loop 594-NWD-596 of the UBA domain has a  
critical role in the interaction with nucleoporins.  
CC -!- DOMAIN: The leucine-rich repeats and the NTF2-domain are  
essential for the export of mRNA from the nucleus.  
CC -!- MISCELLANEOUS: The RNA-binding domain is a non-canonical RNP-type  
domain.  
CC -!- SIMILARITY: BELONGS TO THE NXF FAMILY.  
CC -!- SIMILARITY: CONTAINS 4 LEUCINE-RICH REPEATS (LRR).  
CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
CC -!- SIMILARITY: CONTAINS 1 UBA DOMAIN.

THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; AJ132712; CAA10753.1; -;  
DR EMBL; AF112880; AAD39102.1; -;  
DR EMBL; AF126246; AAD20016.1; -;  
DR EMBL; BC004904; AAH04904.1; -;  
DR EMBL; AK027192; -; NOT ANNOTATED\_CDS.  
DR EMBL; U80073; AAB81111.1; -;

DR Genew; HGNC:8071; NKX1.  
 DR MIM; 602647; -.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003603; LRRcap.  
 DR PDB; 1GO5; 18-OCT-01.  
 DR PDB; 1JRG; 12-JUL-01.  
 DR PDB; 1JNS; 23-JUL-01.  
 DR Pfam; PF00560; LRR; 2.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR SMART; SM00446; LRRcap; 1.  
 DR PROSITE; PS50177; NTF2\_DOMAIN; 1.  
 KW Transport; mRNA transport; Nuclear protein; RNA-binding; Repeat;  
 KW Leucine-rich repeat; Multigene family; 3D-structure.  
 FT DOMAIN 119 198  
 FT REPEAT 266 291  
 FT REPEAT 292 315  
 FT REPEAT 316 343  
 FT REPEAT 344 371  
 FT REPEAT 386 536  
 FT DOMAIN 566 608  
 FT DOMAIN 67 100  
 FT DOMAIN 83 110  
 FT DOMAIN 551 561  
 FT MUTAGEN 306 308  
 FT MUTAGEN 594 594  
 FT MUTAGEN 595 595  
 FT MUTAGEN 617 617  
 FT CONFLICT 119 119  
 FT CONFLICT 256 256  
 FT CONFLICT 256 256  
 SQ SEQUENCE 619 AA; 70182 MW; 338872AADA789FBF CRC64;  
 Query Match 7.0%; Score 86; DB 1; Length 619;  
 Best Local Similarity 25.4%; Pred No. 2.3; Mismatches 32; Indels 89; Gaps 14;  
 Matches 61; Conservative 32; Mismatches 32; Indels 89; Gaps 14;  
 Qy 5 AGDSWGL-----ACLCVLMHLPALNRTGDP-GFGPSIQKTYDIT---RYLHQL 54  
 Db 400 SGRQGLLDAYHDGACCSLSIPFQNPAPRSSLAIEYFKDSNNVKLKDPTRPRLKHTR 459  
 Qy 55 RSLATYNYLGPPEPPNP--PRLGAE--TLPRATVN---LEWRSINDBLR-LTQN 106  
 Db 460 LNVV-AFLNEL--PKTQHDVNSFVVDISAQTSTLLCFSVNGVFKEDYDGRSDSLRAFTRT 516  
 Qy 107 Y---EAYSHLLCYLRG--LNROAAATLRLRSLAHFCTSLQGLLGSIAGVMTLGYPLPOP 161  
 Db 517 FIAPVPSNSGLCIVNDELVRNASEIEIQAF-----MPAP 553  
 Qy 162 LPTGTEPAWAPGPAHSDFLQK-----MDDFWLLKEIQTLW---RSKADFNRLKKGKQVPP 212  
 Db 554 TPSSSPVPTLSPEQEMQLQAFSTQSGMNLWSQKCLQDNNWDYTRSAQAFTHLKAKGEIP 613  
 RESULT 8  
 ID MIPL SCHPO STANDARD; PRT; 1313 AA.  
 AC P87141;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE WD-Repeat protein mipl.  
 GN MIPL OR SPAC57A7.11.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
 EX MEDLINE=20115869; PubMed=10648609;  
 RA Shinozaki-Yabana S., Watanabe Y., Yamamoto M.;  
 RT "Novel WD-repeat protein Miplp facilitates function of the meiotic  
 RT regulator Mei2p in fission yeast.";

RL Mol. Cell. Biol. 20:1234-1242(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Rhode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: BINDS TO AND FACILITATES THE FUNCTIONING OF THE MEIOTIC  
 CC REGULATOR MEI2. MAY ALSO BE INVOLVED IN CONJUGATION BY INTERACTING  
 CC WITH STE11. ESSENTIAL FOR CELL GROWTH.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB032552; BAA84585.1; -.  
 CC EMBL; Z95396; CAB08769.1; -.  
 CC InterPro; IPR001680; WD40.  
 CC InterPro; IPR004083; Yeast176.  
 CC Pfam; PF00400; WD40; 5.  
 CC PRINTS; PR00320; GPROTEINBRPT.  
 CC PRINTS; PR01547; YEAST176DUF.  
 CC SMART; SM00320; WD40; 5.  
 CC PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 CC PROSITE; PS00882; WD\_REPEATS\_2; 2.  
 CC PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Meiosis; WD repeat; Repeat.  
 FT REPEAT 986 1029 WD 1.  
 FT REPEAT 1033 1074 WD 2.  
 FT REPEAT 1087 1126 WD 3.  
 FT REPEAT 1130 1170 WD 4.  
 FT REPEAT 1176 1216 WD 5.  
 FT REPEAT 1219 1259 WD 6.  
 FT REPEAT 1268 1308 WD 7.  
 SQ SEQUENCE 1313 AA; 148533 MW; C71B663B0171E7A4 CRC64;  
 Query Match 7.0%; Score 86; DB 1; Length 1313;  
 Best Local Similarity 22.2%; Pred. No. 5.9;  
 Matches 58; Conservative 31; Mismatches 84; Indels 88; Gaps 12;  
 Qy 15 ICTVLMHLPV--PALNR-----TGDGPGPSI-----OKTYDLTRYLHQLRS 56  
 Db 811 LAFLLQHLPALHKASLKSDTNTSVTSDFKPHFPVPSVSENKILNRSLTRSLKGLAUS 870



```
QY 57 LAG-----TYLNYL-----GPP-FNEPDPNPPRLGAETLPRATVNL 91
   |||
   :|||:|||||:|:
Db 871 LAGDRASELLSNGENKPAESNLNHLTSAKVPGPAFNELEYQ-----SELDMLTSLYLF 926
   :|||:|||||:|:
QY 92 EVWRLNDRRLRLTONYAYSHLLCYLRL-----NRQAATAELRRSLAHFCTS-----LOG 142
   :|||:|||||:|:
Db 927 DWSRKYFTEPQWRNEDDEPGICYNORLWRNRNREKLIYRTRPLAEYSTNGRWNOQLMT 986
   :|||:|||||:|:
QY 143 LLGSIAGVWATLGYPPLQPLPOTEPAPGPAHSDFLQKMDDFWLKKE---LQTLWLRSA 199
   :|||:|||||:|:
Db 987 FNNVIA-----PRKLMFHOFEDQLITLGDKDIIOVMDWRN 1022
   :|||:|||||:|:
QY 200 KDFNRLKKMQPPAASVTLHL 220
   :|||:|||||:|:
Db 1023 RCLNSFK-----TSASATTV 1038
   :|||:|||||:|:

RESULT 9
PHAC_PSEOL STANDARD; PRT; 560 AA.
AC P26496;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Poly(3-hydroxyalkanoate) polymerase 2 (BC 2.3.1.-) (PHA polymerase 2)
DE (PHA synthase 2) (Polyhydroxyalkanoic acid synthase 2).
GN PHAC.
OS Pseudomonas oleovorans.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=301;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GP01;
RX MEDLINE=91115830; PubMed=1989978;
RA Huisman G.W., Wouink E., Meima R., Kazemier B., Terpstra P.,
RA Witholt B.;
RT "Metabolism of poly(3-hydroxyalkanoates) (PHAs) by Pseudomonas
RT oleovorans. Identification and sequences of genes and function of the
RT encoded proteins in the synthesis and degradation of PHA.";
RL J. Biol. Chem. 266:2191-2198(1991).
CC -1- FUNCTION: P.OLEOVORANS ACCUMULATES POLY(3-HYDROXYALKANOATES) AFTER
CC GROWTH ON MEDIUM CHAIN LENGTH HYDROCARBONS. LARGE AMOUNTS OF THIS
CC POLYESTER ARE SYNTHESIZED WHEN CELLS ARE GROWN UNDER NITROGEN-
CC LIMITING CONDITIONS. WHEN NITROGEN IS RESUPPLIED IN THE MEDIUM,
CC THE ACCUMULATED PHA IS DEGRADED.
CC -1- SIMILARITY: BELONGS TO THE PHA/PHB SYNTHASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M58445; AAA25934.1; -.
DR PIR; C38604; C38604.
DR InterPro; IPR000073; Abhydrolase.
DR Pfam; PF00561; abhydrolase; 1.
KW PHA biosynthesis; Transferase; Acyltransferase.
FT ACT SITE 296 296 POTENTIAL.
SQ SEQUENCE 560 AA; 62631 MW; EACD844FC1616B83 CRC64;

Query Match 7.0%; Score 85.5; DB 1; Length 560;
Best Local Similarity 30.2%; Pred. No. 2.3;
Matches 38; Conservative 16; Mismatches 43; Indels 29; Gaps 7;

QY 77 PRLGAETLPRATVNLVWRSNDRRLRLTONYAYSHLLCYLRLGRLNQAATAELRRSL--A 134
   |||
   :|||:|||||:|:
Db 5 PAKGTPTLPATSMNVQ-----NAILGLRGR-----DLISLRLNRSRQS-----LRHPLHTA 50
   :|||:|||||:|:
```

```
QY 135 HFCTSLQGLGSIAGVWATLGYPPLQPLPG-----TEPAWAPGPAHSDFLQKMDDFWLKKE 190
   :|||:|||||:|:
Db 51 HHLLAGGQLGRV-----ILGDTPLQNPDRPDSPTWSQNPYRRGLQA-----YLAWQ 101
   :|||:|||||:|:
QY 191 LQTLW 196
   |||
   :|||:|||||:|:
Db 102 KQTRLW 107
   :|||:|||||:|:

RESULT 10
HUPK_RHOCA STANDARD; PRT; 294 AA.
AC P30797;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrogenase expression/formation protein hupK.
GN HUPK.
OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91117833; PubMed=2007559;
RA Xu H.W., Wall J.D.;
RT "Clustering of genes necessary for hydrogen oxidation in Rhodobacter
RT capsulatus.";
RT J. Bacteriol. 173:2401-2405(1991).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33303 / B10;
RX MEDLINE=93268090; PubMed=8497190;
RA Colbeau A., Richard P., Toussaint B., Caballero F.J., Elster C.,
RA Delphin C., Smith R.L., Chabert J., Vignais P.M.;
RT "Organization of the genes necessary for hydrogenase expression in
RT Rhodobacter capsulatus. Sequence analysis and identification of two
RT hyp regulatory mutants.";
RL Mol. Microbiol. 8:15-29(1993).
CC -1- SIMILARITY: BELONGS TO THE HUPK FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M55089; AAA72924.1; -.
DR EMBL; Z15089; CAA78803.1; -.
DR PIR; B38532; B38532.
DR PIR; B38532; B38532.
FT CONFLICT 70 70 A -> R (IN REF. 2).
SQ SEQUENCE 294 AA; 30222 MW; 2669E8B70AF35EDA CRC64;

Query Match 6.9%; Score 84.5; DB 1; Length 294;
Best Local Similarity 31.2%; Pred. No. 1.2;
Matches 40; Conservative 3; Mismatches 38; Indels 47; Gaps 6;

QY 76 PPRLGAETLPRATVNLVWRSNDRRLRLTONYAYSHLLCYLRLGRLNQAATAE---LRRS 132
   |||
   :|||:|||||:|:
Db 36 PFOQVALLPR-----LNLGCAAGHARLAL-----GLPAEAPARREILRDH 80
   :|||:|||||:|:
QY 133 LAHFCTSLQGLGSIAGVWATLGYPPLQPLPOTEPAPGPAHSDFLQKMDDFWLKKELO 192
   |||
   :|||:|||||:|:
Db 81 LAKLCLIPWKLJGU-----APQPLP-----EHWAGGA-----ALQ 111
   :|||:|||||:|:
QY 193 TWLWRSK 200
   |||
   :|||:|||||:|:
Db 112 HNLWGGAK 119
   :|||:|||||:|:
```

RESULT 11  
M2B2\_PIG STANDARD; PRT; 995 AA.  
AC Q28949;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Epididymis-specific alpha-mannosidase precursor (EC 3.2.1.24)  
DE (Mannosidase alpha class 2B member 2) (AMAN).  
GN MAN2B2.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX TISSUE=epididymis; PubMed=8562059;  
RA Okamura N., Tamba M., Liao H.-J., Onoe S., Sugita Y., Dacheux F.,  
Dacheux J.-L.;  
RT "Cloning of complementary DNA encoding a 135-kilodalton protein  
secreted from porcine corpus epididymis and its identification as an  
epididymis-specific alpha-mannosidase.";  
RL Mol. Reprod. Dev. 42:141-148(1995).  
CC -!- FUNCTION: CAN DIGEST BOTH P-NITRO-PHENYL-ALPHA-D-MANNOSIDE AND  
HIGH MANNOSE OLIGOSACCHARIDE (MAN(8)-GLCNAC(2)). MAY BE INVOLVED  
IN SPERM MATURATION. HAS A POSSIBLE ROLE IN SPECIFIC SPERM-BGG  
INTERACTION SINCE SPERM SURFACE MANNOSIDASE ACTS LIKE A RECEPTOR  
FOR MANNOSE-CONTAINING OLIGOSACCHARIDES LOCATED ON THE ZONA  
PELLUCIDA.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-  
mannose residues in alpha-D-mannosides.  
CC -!- SUBCELLULAR LOCATION: SECRETED. FOUND AT THE SPERM SURFACE AS A  
27 kDa FRAGMENT.  
CC -!- TISSUE SPECIFICITY: SPECIFIC TO THE CAPUT AND CORPUS OF THE  
EPIDIDYMIS.  
CC -!- PTM: PROCESSED INTO A 27 kDa FRAGMENT LOCALIZED ON THE EQUATORIAL  
SEGMENT AND THE APICAL RIM OF THE HEAD OF MATURE SPERM.  
CC -!- MISCELLANEOUS: OPTIMAL MANNOSIDASE ACTIVITY WAS FOUND AT PH 6.5.  
CC -!- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; D28521; BAA05877.1; ALT INIT.  
DR InterPro; IPR000602; Glyco\_hydro\_38.  
PFam; PF01074; Glyco\_hydro\_38; 2\_  
KW Hydrolase; Glycosidase; Signal; Glycoprotein.  
FT SIGNAL  
FT CHAIN 1 21  
FT 22 995 EPIDIDYMIS-SPECIFIC ALPHA-MANNOSIDASE.  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 593 593 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 625 625 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 657 657 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 733 733 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 793 793 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 875 875 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 977 977 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 945 945 D -> A (IN REF. 1; AA SEQUENCE).  
FT SEQUENCE 995 AA; 114230 MW; 79D21B8CE5AE6FC2 CRC64;  
Query Match 6.6%; Score 80.5; DB 1; Length 995;  
Best Local Similarity 23.0%; Pred. No. 14;  
Matches 45; Conservative 22; Mismatches 62; Indels 67; Gaps 9;  
QY 50 LEHQ-----LRSAGYLYNLVGLPPFPNPPRLGAEYL-----PRATVLEWRSL 97  
DB 827 LOHRPVLFRLEGGTVQNGPGRKQEPVTLPPSLHLQILSPGWKYSNHTVHLKNLQKG 886

QY 98 NDR-----LRLTONYEAYSHLLCYLRGLNRQATAELRRSLAHFCTSLQGLLSI 147  
DB 887 HYERRAKADFERVLLRLHLLYEAGEH-----QALSRPVT-----LNLSQSVLRGL 929  
QY 148 AGVMATLGVPLPQPLPGTEPAMAPGPAHSDFLQKMDDFWLKELQTLWLRSAKPNR--L 205  
DB 930 GSVVAV-----EERSLTGT-----WDVNSLHRWSWKTEDGHHRRGS 965  
QY 206 KKKMQPP--AASVTLH 219  
DB 966 SRRPLPLRGPNTVTH 981  
RESULT 12  
VPP3\_HUMAN STANDARD; PRT; 830 AA.  
AC Q13488; O75877;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Vacuolar proton translocating ATPase 116 kDa subunit A isoform 3 (V-  
ATPase 116-kDa isoform a3) (Osteoclastic proton pump 116 kDa subunit)  
DE (OC-116 kDa) (OC116) (T-cell immune regulator 1) (T cell immune  
response cDNA7 protein) (TIRC7).  
GN TCIRG1 OR ATP6N1C.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RC TISSUE=Osteoclastoma;  
RX MEDLINE=96158968; PubMed=8579597;  
RA Li Y.P., Chen W., Stashenko P.;  
RT "Molecular cloning and characterization of a putative novel human  
osteoclast-specific 116-kDa vacuolar proton pump subunit.";  
RL Biochem. Biophys. Res. Commun. 218:813-821(1996).  
RN [2]  
RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
RA Utku N., Heinemann T., Bulwin C.-G., Beinke S., Beato F., Randall J.,  
Busconi L., Delphire E., Robertson E.R., Kojima R., Volk H.D.,  
Milford E.L., Gullans S.R.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
RX MEDLINE=99263502; PubMed=10329006;  
RA Heinemann T., Bulwin G.C., Randall J., Schnieders B., Sandhoff K.,  
Volk H.D., Milford E., Gullans S.R., Utku N.;  
RT "Genomic organization of the gene coding for TIRC7, a novel membrane  
protein essential for T cell activation.";  
RL Genomics 57:398-406(1999).  
CC -!- FUNCTION: PART OF THE PROTON CHANNEL OF V-ATPASES (BY SIMILARITY).  
CC SEEMS TO BE DIRECTLY INVOLVED IN T CELL ACTIVATION.  
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A,  
C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE  
COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: THE LONG ISOFORM IS HIGHLY EXPRESSED IN  
OSTEOCLASTOMAS. THE SHORT FORM IS HIGHLY EXPRESSED IN THYMUS.  
CC -!- SIMILARITY: BELONGS TO THE V-ATPASE 116 kDa SUBUNIT FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----



large proline-rich proteins with multiple repeated motifs and a single ubiquitin-like domain.";  
 Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).  
 CC -!- FUNCTION: UNKNOWN.  
 CC -!- SIMILARITY: CONTAINS 1 UBIQUITIN-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M33519; AAA35587.1; -.  
 CC EMBL; M33521; AAA35588.1; -.  
 CC EMBL; M33520; AAA35588.1; JOINED.  
 CC HSPF; P02248; IUBI.  
 CC Genew; HGNC:13919; BAT3.  
 CC MIM; 142590; -.  
 CC InterPro; IPR000626; Ubiquitin.  
 CC Pfam; PF00240; ubiquitin; 1.  
 CC SMART; SM00213; UBQ; 1.  
 CC PROSITE; PS00299; UBIQUITIN\_1; 1.  
 CC PROSITE; PS00553; UBIQUITIN\_2; 1.  
 KW Repeat.  
 FT DOMAIN 17 77 UBIQUITIN-LIKE.  
 FT DOMAIN 202 207 POLY-PRO.  
 FT DOMAIN 242 636 4 X 29 AA APPROXIMATE REPEATS.  
 FT REPEAT 242 270 1.  
 FT REPEAT 415 443 2.  
 FT REPEAT 474 602 3.  
 FT REPEAT 608 636 4.  
 FT DOMAIN 657 670 POLY-PRO.  
 SQ SEQUENCE 1132 AA; 119504 MW; 528CA8A78C39DD18 CRC64;  
 Query Match 6.5%; Score 79; DB 1; Length 1132;  
 Best Local Similarity 26.1%; Pred. No. 22;  
 Matches 46; Conservative 17; Mismatches 67; Indels 46; Gaps 11;  
 QY 23 PAVPALNRGTGPGPSIKYKTDLYLHQLSLAGTYLNYLPGPFNEDFPNPRIGAE 82  
 Db 261 PA-PETNAPNHPSPAYEVVLEQLQ-LRSRLQPFQRYEVLGAAAT-TDYNNHEGRE 317  
 QY 83 TLPRATVNLVWRSLNDRLLRQTN-YEAYSHLLCYLRGLNRQAT----AELRRSLAHFC 137  
 Db 318 EQDR-LINL-----VGESLRLLGNTFVALSDLRCLN-----ACTPRHLHVVRPKSHYT 365  
 QY 138 TSQGLGSGIAGMATLGYPL-----POPLPQTE-PAWAPGPAHS 176  
 Db 366 TPM-----VLQQAALPIQINVGTTVTMTGNGTRPPTPNAEAPPPGQASS 412  
 RESULT 15  
 ID ALK HUMAN STANDARD; PRT; 1620 AA.  
 AC Q9UM73; Q9Y4K6;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE ALK tyrosine kinase receptor precursor (SC 2.7.1.112) (Anaplastic  
 DE lymphoma kinase) (CD246 antigen).  
 GN ALK.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_taxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND  
 RP GLYCOSYLATION.  
 RX MEDLINE=97316779; PubMed=9174053;  
 RA MORRIS S.W., Naewe C.W., Mathew P., James P.L., Kirstein M.N., Cui X.,  
 RA Witte D.P.;

"ALK, the chromosome 2 gene locus altered by the t(2;5) in non-Hodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK).";  
 Oncogene 14:2175-2188(1997).  
 RL [2]  
 RN ERRATUM.  
 RP MORRIS S.W., Naewe C.W., Mathew P., James P.L., Kirstein M.N., Cui X.,  
 RA Witte D.P.;  
 RA Oncogene 15:2883-2883(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97178863; PubMed=9053841;  
 RA Iwahara T., Fujimoto J., Wen D., Cupples R., Bucay N., Arakawa T.,  
 RA Mori S., Razukin B., Yamamoto T.;  
 RT "Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system.";  
 RL Oncogene 14:439-449(1997).  
 RN [4]  
 RP PARTIAL SEQUENCE FROM N.A., AND CHROMOSOMAL TRANSLOCATION.  
 RX MEDLINE=94167588; PubMed=812112;  
 RA MORRIS S.W., Kirstein M.N., Valentine M.B., Dittmer K.G.,  
 RA Shapiro D.N., Saltman D.L., Look A.T.;  
 RT "Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma.";  
 RL Science 263:1281-1284(1994).  
 CC -!- FUNCTION: Orphan receptor with a tyrosine-protein kinase activity. Appears to play an important role in the normal development and function of the nervous system.  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- TISSUE SPECIFICITY: Expressed in brain and CNS. Also expressed in the small intestine and testis, but not in normal lymphoid cells.  
 CC -!- PTM: N-glycosylated.  
 CC -!- DISEASE: A FORM OF NON-HODGKIN'S LYMPHOMA IS CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(2;5) (P23;Q35) THAT INVOLVES NPM1 AND ALK.  
 CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN RECEPTOR SUBFAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 2 MAM DOMAINS.  
 CC -!- DATABASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.; WWW="http://www.infobiogen.fr/services/chromocancer/Genes/ALK.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U62540; AAB71619.1; -.  
 CC EMBL; U66559; AAC51104.1; -.  
 CC HSPF; P08631; IAD5.  
 CC Genew; HGNC:427; ALK.  
 CC MIM; 105590; -.  
 CC InterPro; IPR000719; Euk\_pkinase.  
 CC InterPro; IPR002172; LDL\_recept\_A.  
 CC InterPro; IPR000998; MAM\_domain.  
 CC InterPro; IPR002011; RTKinaseII.  
 CC InterPro; IPR001245; Tyr\_pkinase.  
 CC Pfam; PF00069; pkinase; 1.  
 CC Pfam; PF00629; MAM; 1.  
 CC PRINTS; PR00109; TYRKINASE.  
 CC ProDom; PD000001; Euk\_pkinase; 1.  
 CC SMART; SM00192; LDLA; 1.  
 CC SMART; SM00219; TyrKc; 1.  
 CC PROSITE; PS00740; MAM\_1; FALSE\_NEG.  
 CC PROSITE; PS50060; MAM\_2; 2.  
 CC PROSITE; PS01209; LDLA\_1; FALSE\_NEG.  
 CC PROSITE; PS50068; LDLA\_2; FALSE\_NEG.  
 CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
DR PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
DR PROSITE; PS00239; RECEPTOR\_TYR\_KIN\_II; 1.  
KW Transferase; Tyrosine-protein kinase; Transmembrane; ATP-binding;  
KW Phosphorylation; Receptor; Glycoprotein; Repeat; Signal;  
KW Proto-oncogene; Chromosomal translocation.  
FT SIGNAL 1 18  
FT CHAIN 19 1620  
FT DOMAIN 19 1038  
FT TRANSFEM 1039 1059  
FT DOMAIN 1060 1620  
FT DOMAIN 264 427  
FT DOMAIN 437 473  
FT DOMAIN 478 636  
FT DOMAIN 1116 1392  
FT DOMAIN 816 940  
FT NP\_BIND 1122 1130  
FT BINDING 1150 1150  
FT ACT\_SITE 1249 1249  
FT MOD\_RES 1282 1282  
FT CARBOHYD 169 169  
FT CARBOHYD 244 244  
FT CARBOHYD 285 285  
FT CARBOHYD 324 324  
FT CARBOHYD 411 411  
FT CARBOHYD 424 424  
FT CARBOHYD 445 445  
FT CARBOHYD 563 563  
FT CARBOHYD 571 571  
FT CARBOHYD 627 627  
FT CARBOHYD 709 709  
FT CARBOHYD 808 808  
FT CARBOHYD 863 863  
FT CARBOHYD 864 864  
FT CARBOHYD 886 886  
FT CARBOHYD 986 986  
FT CONFLICT 36 36  
FT CONFLICT 1491 1491  
FT CONFLICT 1529 1529  
SQ SEQUENCE 1620 AA; 176417 MW; A62604B242961E1E CRC64;  
  
Query Match 6.5%; Score 79; DB 1; Length 1620;  
Best Local Similarity 23.1%; Pred. No. 35;  
Matches 54; Conservative 15; Mismatches 77; Indels 88; Gaps 9;  
  
Qy 18 VLWHLPAVPAL-----NRTGDPGPSIOKTYDLYLEHQLSLAGTYLVNLGP 67  
Db 6 LLWLLPLLSTAAVSGMGWGQAGSPAGSPLPQPREPLS-YSLQKSLA----- 55  
  
Qy 68 PFNEPDPNPRLGAETLPRAVTNLEVMRSLNDRRLTQNYEAYSH-LLCYLRGLNQAA 126  
Db 56 ----VDFVPSL-----FRVYARDLLPPSSSELKAGR 84  
  
Qy 127 AELRRSLAHFTSLQGLGSLAGVNMATLGYPLQPLP----- 163  
Db 85 PEARGSLALDCAPLRLLLGAPAGVSWTAGSPAPAEARTLSRLKGSVRKLRAKQLVLE 144  
  
Qy 164 -GTEP-----AWAPGPAHSDFLQKDDFWLLKELQTLWRSKADFNLKXMQP 211  
Db 145 LGEAILEGCVGPPGGAAGVLLQ-----FNLSELFSWIRQGE--GRURILMP 191

Search completed: January 27, 2003, 15:42:00

Job time : 11 secs





```
|||||
Db 121 NRQAATAEIIRSLAHFCTSLQGLGSIAGVMTLGYPLPQPLGTEPAWAPGAHSDFLQ 180
QY 181 KMDDFWLLKELQTLWRSKADFNRLKKMKQPPAASVTLHLEAHGF 225
Db 181 KMDDFWLLKELQTLWRSKADFNRLKKMKQPPAASVTLHLEAHGF 225

RESULT 2
Q9UBD9 PRELIMINARY; PRT; 225 AA.
AC Q9UBD9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Neurotrophin-1/B-cell stimulating factor-3 (Cardiotrophin-like
DE cytokine) (Similar to cardiotrophin-like cytokine,
DE neurotrophin-1/B-cell stimulating factor-3).
GN CLC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432254; PubMed=10500198;
RA Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S.,
RA Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,
RA Simonet W.S., Boone T., Chang M.-S.;
RT "Novel neurotrophin-1/B cell-stimulating factor-3: A cytokine of the
RT IL-6 family.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99382254; PubMed=10448081;
RA Shi Y., Wang W., Yourey P.A., Gohari S., Zukauskas D., Zhang J.,
RA Ruben S., Alderson R.F.;
RT "Computational EST database analysis identifies a novel member of the
RT neurotrophic cytokine family.";
RL Biochem. Biophys. Res. Commun. 262:132-138(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX Hu X., Xu Y., Zhang B., Peng X., Yuan J., Qiang B.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX Strausberg R.;
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176912; AAF00992.1; -
DR EMBL; AF172854; AAD54284.1; -
DR EMBL; AF176911; AAF00991.1; -
DR EMBL; AY049779; AAL15436.1; -
DR EMBL; BC012939; AAH12939.1; -
SQ SEQUENCE 225 AA; 25176 MW; E2DD4B6280833B55 CRC64;

Query Match 97.5%; Score 1193; DB 4; Length 225;
Best Local Similarity 96.3%; Pred. No. 2.6e-103;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGMGLACTVLMHLPVAPALNRTGDPGPGSIQKTYDLYLRYLHQLRSLAGT 60
Db 1 MDLRAGDSWGMGLACTVLMHLPVAPALNRTGDPGPGSIQKTYDLYLRYLHQLRSLAGT 60
QY 61 YLNYLGPPFPNPPRLGAETLPRATVNLVWRSNDRRLTQNTYEAYSHLCYLRLGL 120
Db 61 YLNYLGPPFPNPPRLGAETLPRATVNLVWRSNDRRLTQNTYEAYSHLCYLRLGL 120
QY 121 NRQAATAEIIRSLAHFCTSLQGLGSIAGVMTLGYPLPQPLGTEPAWAPGAHSDFLQ 180
Db 121 NRQAATAEIIRSLAHFCTSLQGLGSIAGVMTLGYPLPQPLGTEPAWAPGAHSDFLQ 180
QY 181 KMDDFWLLKELQTLWRSKADFNRLKKMKQPPAASVTLHLEAHGF 225
```

```
|||||
Db 181 KMDDFWLLKELQTLWRSKADFNRLKKMKQPPAASVTLHLEAHGF 225

RESULT 3
Q9PUJ2 PRELIMINARY; PRT; 215 AA.
AC Q9PUJ2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Receptivity factor isoform 1 precursor.
GN PRF.
OS Plethodon jordani (Salamander).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Plethodontidae;
OC Plethodon.
OX NCBI_TaxID=8336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420364; PubMed=10489368;
RA Rollmann S.M., Houck L.D., Feldhoff R.C.;
RT "Proteinaceous pheromone affecting female receptivity in a terrestrial
RT salamander.";
RL Science 285:1907-1909(1999).
RW EMBL; AF181480; AAF01025.1; -
KW Signal.
FT SIGNAL.
SQ SEQUENCE 215 AA; 24138 MW; B1906BB666335738 CRC64;

Query Match 11.8%; Score 144.5; DB 13; Length 215;
Best Local Similarity 25.6%; Pred. No. 1.3e-05;
Matches 41; Conservative 31; Mismatches 81; Indels 7; Gaps 3;

QY 56 SLACTYNYLGPPFPNPPRLGAETLPRATVNLVWRSNDRRLTQNTYEAYSHLC 115
Db 55 SLLPTLYSLFQAGAPISDDPYQLPHIKVANLPTAAMDYDTFMKQTDTRLNNNYFYSAYIE 114
QY 116 YLR-GLNRQ----AATAEIIRSLAHFCTSLQGLGSIAGVMTLGYPLPQPLGTEPAWA 170
Db 115 FLKEAMTEQEDLNPAELSLKAKFEAMANSNTLSKISDINTQMGMSVTITLP--KPLVV 172
QY 171 PGPAHSDFLQMDDFWLLKELQTLWRSKADFNRLKKMKQ 210
Db 173 PFECSAVPRKKLGGVCKEYKERVLLTKRDFEFLAKKYQ 212

RESULT 4
Q9PUJ1 PRELIMINARY; PRT; 215 AA.
AC Q9PUJ1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Receptivity factor isoform 2 precursor.
GN PRF.
OS Plethodon jordani (Salamander).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Plethodontidae;
OC Plethodon.
OX NCBI_TaxID=8336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420364; PubMed=10489368;
RA Rollmann S.M., Houck L.D., Feldhoff R.C.;
RT "Proteinaceous pheromone affecting female receptivity in a terrestrial
RT salamander.";
RL Science 285:1907-1909(1999).
RW EMBL; AF181481; AAF01026.1; -
KW Signal.
FT SIGNAL.
SQ SEQUENCE 215 AA; 24080 MW; B341B8B7B4E28438 CRC64;
```





```
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
DR EMBL; AL021409; CAA16181.1; -.
DR HSSP; P12998; 1BS0.
DR InterPro; IPR004839; AminoTransf1/2.
DR InterPro; IPR003880; Pnante. attach.
DR Pfam; PF00155; aminotran_1_2; 1.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS00075; ACP_DOMAIN; 1.
KW Phosphopantetheine.
SQ SEQUENCE 640 AA; 68061 MW; E619FBB5B55A613F CRC64;

Query Match 7.6%; Score 93.5; DB 16; Length 640;
Best Local Similarity 28.3%; Pred. NO. 2.9;
Matches 52; Conservative 12; Mismatches 63; Indels 57; Gaps 9;

QY 22 LPAPVAPALNRGTGDPGPGPSIQKTYDLTRYLEHQL---RSLAGTYLNYLGGPPFNEPDENPPR 78
DB 100 LPAPVA-EGTGDPTVEAVAAAMRQTQVRHQLGLDADLEG-----GAAPDAAGATTTLRALA---D 179

QY 79 LGAETLPATVNLVWRSINLRLRLTONYAYSHLLCYLRLGNLRQAATAEALRRSLAHFCT 138
DB 141 LGVDSVLTLSVVAEA---TERLGLT-----GAAPDAAGATTTLRALA---D 179

QY 139 SLOGLLGSIAGWMATLGYPLPGTEPAWAPGPAHSDFLQKMDDFWLLKELQTLWLS 198
DB 180 ALRGL-----VAAAPGTAVPAPATGAA-APAPGSGNAP-----APGADGWDHRS 225

QY 199 AKDF 202
DB 226 MKDF 229

RESULT 8
O88841 PRELIMINARY; PRT; 727 AA.
AC O88841;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Faciogenital dysplasia protein 2.
GN FGD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RA MEDLINE=93389726; PubMed=10458911;
RA Pastoris N.G.; Gorski J.L.;
RT "Isolation, characterization, and mapping of the mouse and human fgd2
RT genes, faciogenital dysplasia (FGD1; aarskog syndrome) gene
RT homologues."
RL Genomics 60:57-66(1999).
DR EMBL; AF017368; AAC35430.1; -.
DR MGB; MGI:1347084; FGD2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM0064; FYVE; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
SQ SEQUENCE 727 AA; 82039 MW; 5CD0FBFB905C5FC5 CRC64;

Query Match 7.6%; Score 93; DB 11; Length 727;
Best Local Similarity 24.8%; Pred. NO. 3.9;
Matches 61; Conservative 32; Mismatches 93; Indels 60; Gaps 11;
```

```
QY 21 HLPAPVAPALNRGTGDPGPGPSIQKTYDLTRYLEHQL-----RSLAGT 60
DB 13 NLVAVFENNRT--PGEAPGSHSLEDQPHIPEHQLSLSPPEWAPPVKEALKSEPRPVSR 70
QY 61 YLNYLGGPPFN-----BPDFNPPLGATLPRATVNLVWRSINLRLRLTONYAYSH 112
DB 71 YLSLKNKLSSGAWRRSCQPGVSP--GPE-----TOPEBEKRVVRELLETEQAYVARLH 122
QY 113 LL--CYLRLGNLRQAATAEALRRSLAHFCTSLQGLLGSIAGWMA-TLGYPLPGTEPAW 169
DB 123 LLDQVFFQELLREAG-----RSKAFEDVVVKLIFSNLISSYRFAHQFLLP-LORRVD 176
QY 170 ACPAHSDFLQKMDDFWLL-----KEIQTWLSAKDFNRLKKMQPPAASVT 217
DB 177 AATPRIGDVIQKLAPFLKMYSEYVKNPFAAEALLATWMDKS-QPFQEVTVTIQSEASS 235
QY 218 LHLEAH 223
DB 236 LTLQHH 241

RESULT 9
O8YBB1 PRELIMINARY; PRT; 200 AA.
AC O8YBB1;
DT 01-WAR-2002 (TrEMBLrel. 20, Created)
DT 01-WAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein BMEII0989.
GN BMEII0989.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
[1]
SEQUENCE FROM N.A.
RA STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RA MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G.; Kapral V.; Redkar R.J.; Patra G.; Mujar C.; Los T.;
RA Ivanova N.; Anderson I.; Bhattacharya A.; Lykidis A.; Reznik G.;
RA Jablonski L.; Larsen N.; D'Souza M.; Bernal A.; Mazur M.; Goltzman E.;
RA Selkov E.; Elzer P.H.; Hagius S.; O'Callaghan D.; Letesson J.-J.;
RA Haselkorn R.; Kyripides N.; Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009732; AAL54231.1; -.
DR InterPro; IPR00734; Lipase.
DR PROSITE; PS00120; LIPASE_SER; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 200 AA; 21513 MW; B4A8596707B67617 CRC64;

Query Match 7.6%; Score 92.5; DB 16; Length 200;
Best Local Similarity 23.9%; Pred. NO. 0.82;
Matches 55; Conservative 21; Mismatches 89; Indels 65; Gaps 11;

QY 10 GMLACLTVLWHLPAVAPALNRGTGDPGPGPSIQKTYDLTRYLEHQLRSLAGTYLNYLGGPP 69
DB 2 GLLAGAALV---LPSLPAEAKTQQAAMPFNATSPHQADVYL---LRGFADIFSTGI--- 51
QY 70 NEPDENPRLCAETLPATVNLV-----WRSINLRLRLTONYAYSHLLCYLRLGNRQA 124
DB 52 -----DEIGAE-LQAAGVNAHVQHAWEVLNRIADVQKNGHLFVVLIGHSLGANA 103
QY 125 A---TAEALRRSLAHFCTSLQGLLGSIAGWMATLGYPLPGTEPAW-----EPAW-- 169
DB 104 AIYIAEELERR-----GIAVDYMATFAATGDPPLPGNVRRVNVNFFYKQHGWL 151
QY 170 --APGPAHSDFLQKMDDFWLLKELQTLWLSAKDFNRLK-KMQPPAASV 216
DB 152 PLVGPGRFHGHLENRD-----FSNAKGVGHFNIEKQRFLOAEV 189
```

```

RESULT 10
Q9MAU1 ID Q9MAU1 PRELIMINARY; PRT; 332 AA.
AC Q9MAU1
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F13M7.8 protein.
GN F13M7.8
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP "The sequence of BAC F13M7 from Arabidopsis thaliana chromosomes 1.";
RT Submitted (JUN-1998) to the EMBL/GenBank/DBSJ databases.
RL
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Kwan A., Liu, S.,
RA Li J., Kremenetskaia I., Luros J., Araujo R., Au M., Bredel V.,
RA Buehler E., Conway A., Dewar K., Feng J., Kim C., Kurtz D., Li Y.,
RA Palm C., Shinn P., Sun H., Davis R., Ecker J., Federspiel N.,
RA Theologis A.;
RT "The sequence of BAC F13M7 from Arabidopsis thaliana chromosomes 1.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBSJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL EMBL; AC004809; AAF40444.1; -.
DR InterPro; IPR002965; P-rich extensin.
DR PRINTS; PR01217; PRICEXTENSIN.
SQ SEQUENCE 332 AA; 36793 MW; 18E8687141A070F4 CRC64;

Query Match 7.5%; Score 92; DB 10; Length 332;
Best Local Similarity 25.9%; Pred. No. 1.8;
Matches 57; Conservative 24; Mismatches 73; Indels 66; Gaps 12;

QY 23 PAVPALNRGTGPGGPGSIQKTYDLYLRYLHQLRSAGTYLNY----- 64
Db 113 PSVTAGNLGYP-PRPSP--TYDPGYBQRMESLLQQFIRNRPQIRPLRLGLGSPVG 169
QY 65 LGPFNEPDPNPRGLGAETLPRATNVNLEVRSLNDRRLRTQNYEAYSHLLCYLRLGNRQA 124
Db 170 LGPIRASPFQLQPRVA-----PPPTSLDTSNRKARSK-----DGLAVVRG--RKV 215
QY 125 ATAELRRL-----AHFCTSLQGLLSIAGVMTATLGYPLPQLP--GTEPAWA 170
Db 216 RITGSSSLYSLGRLWLNKAHV-----GIQPRSGIMK-----PLPKPLPVLDTTTSVP 266
QY 171 PGPAHSDFLQMDDFWLLKELQTLWRSKADFNRLKKMQ 210
Db 267 DDPDESADEKDEAVKQL-----SEKDL--LKRHIE 298

RESULT 11
Q9CWV7 ID Q9CWV7 PRELIMINARY; PRT; 455 AA.
AC Q9CWV7
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 2410003H12RIK protein.
GN 2410003H12RIK
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
RL MEDLINE=21085660; PubMed=11217851;

```

```

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Komno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauer P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010358; BAB26878.1; -.
DR MGD; MGI:1919221; 2410003H12RIK.
SQ SEQUENCE 455 AA; 51481 MW; 201886B814EB9CFF CRC64;

Query Match 7.5%; Score 91.5; DB 11; Length 455;
Best Local Similarity 23.6%; Pred. No. 2.9;
Matches 59; Conservative 21; Mismatches 75; Indels 95; Gaps 14;

QY 7 DSKGMLACLCTVLW-HLPAYPALNRGTGPG-----PPSIQKTYDLYLRYLHQLRS 56
Db 29 DAFNYOSCFMODLFAHPPEVLFIHRTYTPRGKVLTYFLVDGFRVQVEGLARAV----- 82
QY 57 LAGTYLNYLGGPFNE-----PDFNPPRLGAETLPRATNVNLEVRSLNDRRLRT 104
Db 83 -----YFAIPTNEDARGLAQMFQVFKFNP-----AWERVNTIL----- 116
QY 105 QNYEAYSHLLCYLRLGNRQAATAELRSLAHFCTSLQG-----LLGSIAGV 150
Db 117 ----VDPHFL-LLPTLTMEPTAEVLLSAFHCKFLQKGYQLPLEQPVQVRLLSLQST 171
QY 151 M--ATLG-----YPLPQLPGTEPAWAPGPAHSDFLQMDDFWLLKELQTLW----WRSNAK 200
Db 172 MCSATAGNLRKLYTL---LNNCIP-----SSRLPELHSHWLLND-RIMLAHWRSGRA 219
QY 201 DFNRLKKKMQ 210
Db 220 QSSRYFQSLQ 229

RESULT 12
Q9JHE4 ID Q9JHE4 PRELIMINARY; PRT; 423 AA.
AC Q9JHE4
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Cerebroside sulfotransferase.
GN GCST OR CST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=20193614; PubMed=10727929;
RA Hirahara Y., Tsuda Y., Wada Y., Honke K.;
RT "cDNA cloning, genomic cloning, and tissue-specific regulation of
RT mouse cerebroside sulfotransferase.";
RL Eur. J. Biochem. 267:1909-1917(2000).
DR EMBL; AB032940; BAA93009.1; -.
DR EMBL; AB032939; BAA93008.1; -.
DR MGD; MGI:1858277; Gcat.
RW Transferrase.

```



Search completed: January 27, 2003, 15:44:09  
Job time : 28.5 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:17 ; Search time 33 Seconds  
(without alignments)

908.527 Million cell updates/sec

Title: US-09-931-704-5

Perfect score: 1224

Sequence: 1 MDLRAGDSWGLACTLVLM.....KKKMQPPAASVTLLEAHGF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*  
1: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgcdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1224	100.0	225	19 AAW29716	Mouse neurotrophic
2	1224	100.0	225	19 AAW56142	Amino acid sequenc
3	1224	100.0	225	21 AA187814	Murine NNT-1 prote
4	1224	100.0	225	23 AAU78177	Mouse novel neurot
5	1193	97.5	225	19 AAW29715	Human neurotrophic
6	1193	97.5	225	19 AAW56141	Amino acid sequenc
7	1193	97.5	225	20 AAW94466	Human cardiocroph
8	1193	97.5	225	21 AA187813	Human NNT-1 protei
9	1193	97.5	225	22 AAG63543	Amino acid sequenc
10	1193	97.5	225	23 AAU78176	Human novel neurot

11	1193	97.5	253	22 AAM25831	Human protein sequ
12	1181	96.5	223	22 AAB00828	Human cardiocroph
13	1171	95.7	321	22 ABB11896	Human cardiocroph
14	1171	95.7	321	22 AAM79399	Human protein SEQ
15	1167	95.3	215	21 AAB19587	Mouse interleukin-
16	1164	95.1	260	22 AAM78415	Human protein SEQ
17	1136	92.8	215	21 AAB19586	Human interleukin-
18	852	69.6	164	22 ABB40317	Peptide #7823 enco
19	852	69.6	164	22 ABB24716	Protein #6715 enco
20	852	69.6	164	22 AAM61118	Human brain expres
21	852	69.6	164	22 AAM73827	Human bone marrow
22	852	69.6	164	22 AAM20115	Peptide #6549 enco
23	852	69.6	164	22 AAM34012	Peptide #8049 enco
24	852	69.6	164	23 AAG43716	Human peptide enco
25	156.5	12.8	208	20 AAY09197	Human DNAX interle
26	154	12.6	208	20 AAY09196	Human DNAX interle
27	110.5	9.0	203	16 AAR83965	Mouse cardiac hype
28	110.5	9.0	203	17 AAR88204	Human cardiocroph
29	110.5	9.0	203	18 AAM29237	Murine cardiocroph
30	92.5	7.6	243	22 AAU09153	Human cytokine Zal
31	92.5	7.6	243	22 AAB20277	Human interleukin
32	92.5	7.6	243	23 AAU76375	Human helical prot
33	92	7.5	332	21 AAG22132	Arabidopsis thalia
34	92	7.5	332	21 AAG40321	Arabidopsis thalia
35	91.5	7.5	201	16 AAR83967	Human cardiac hype
36	91.5	7.5	201	18 AAM29238	Human cardiocroph
37	91.5	7.5	201	20 AAY06490	Human tumour-assoc
38	91.5	7.5	201	21 AAB27662	Human protein PRO8
39	91.5	7.5	201	21 AAB13004	Human cardiocroph
40	91.5	7.5	201	21 AAY93697	Amino acid sequenc
41	91.5	7.5	201	21 AAY87818	Human cardiocroph
42	91.5	7.5	201	22 AAB50994	Human PRO82 prote
43	90	7.4	195	14 AAR34432	Sequence of growth
44	90	7.4	195	20 AAW83337	Chicken ciliary ne
45	87	7.1	242	22 AAB20275	Human interleukin

#### ALIGNMENTS

##### RESULT 1

AAW29716  
ID AAW29716 standard; Protein; 225 AA.

XX AAW29716;

AC AAW29716;

DT 09-NOV-1998 (first entry)

XX Mouse neurotrophic factor NNT-1.

XX NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
KW peripheral neuropathy; dystrophy; neural retina degeneration;  
KW common variable immunodeficiency; CVID; selective IgA deficiency;  
KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;  
therapy.

XX Mus sp.

XX Key Location/Qualifiers  
FH Peptide 1..27  
FT Peptide /label= Sig\_peptide

FT Protein 28..225  
FT Protein /label= Mat\_protein

XX WO9833922-A1.

XX 06-AUG-1998.

XX 02-FEB-1998; 98WO-US02363.

XX 30-JAN-1998; 98US-0016534.

PR 03-FEB-1997; 97US-0792019.  
 PA (AMGE-) AMGEN INC.  
 PI Chang M, Elliot GS, Sarmiento U, Senaldi G;  
 XX WPI; 1998-437475/37.  
 DR N-PSDB; AAV47512.  
 XX  
 XX Newly isolated nucleic acid encoding human or murine neurotrophic  
 PT factor NNT-1, useful for treatment of neurological and  
 PT immunological diseases or inflammation, also as vaccine adjuvant  
 XX  
 PS Claim 13; Fig 5; 120pp; English.  
 XX

This is the amino acid sequence of a murine neurotrophic factor,  
 designated NNT-1, that is a growth factor for neurons and for B or  
 T cells. It was deduced from isolated NNT-1 cDNA (see AAV47512).  
 Human NNT-1 (see AAW29715) is also provided. Vectors and host cells  
 for use in the production of human murine recombinant NNT-1  
 polypeptides. These are used to treat: (i) neurological or  
 immunological diseases, specifically Alzheimer's, Parkinson's  
 or Huntington's diseases, amyotrophic lateral sclerosis,  
 Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and  
 degeneration of the neural retina, or conditions characterised by T  
 or B cell defects, e.g. common variable immunodeficiency (CVID),  
 selective IgA deficiency, hypogammaglobulinaemia and X-linked  
 agammaglobulinaemia (claimed), but many others disclosed; and (ii)  
 inflammation. NNT-1 is also able to boost immunoreactivity and  
 antibody production following vaccination, and, since it inhibits  
 tumour necrosis factor production, it may also be useful for  
 treating sepsis. In addition, cells that have been engineered to  
 express NNT-1 can be implanted, or nucleic acids are delivered in  
 gene therapy vectors.

SQ Sequence 225 AA;

Query Match 100.0%; Score 1224; DB 19; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-121;  
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGMACLTVMHLPAVPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60  
 Db 1 MDLRAGDSWGMACLTVMHLPAVPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60  
 QY 61 YLYNLGPPNEPDPNPRLGAEITLPRATVNLVWRSINDRLRLTONYEAYSHLLCYLRGL 120  
 Db 61 YLYNLGPPNEPDPNPRLGAEITLPRATVNLVWRSINDRLRLTONYEAYSHLLCYLRGL 120  
 QY 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180  
 Db 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180  
 QY 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPPAASVTILHLEAHGF 225  
 Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPPAASVTILHLEAHGF 225

RESULT 2  
 AAW56142  
 ID AAW56142 standard; Protein; 225 AA.  
 XX  
 AC AAW56142;  
 XX

DT 13-JUL-1998 (first entry)  
 XX  
 XX Amino acid sequence of murine neurotrophic factor NNT-1.  
 DE

XX Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;  
 KW treatment; neurological disease; degeneration; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.  
 XX  
 OS Mus sp. \*

XX Key Location/Qualifiers  
 FH Peptide 1..27  
 FT /note= "signal peptide"  
 FT Protein 28..225  
 FT /note= "mature peptide"  
 XX  
 PN US5741772-A.  
 XX  
 XX 21-APR-1998.  
 PD  
 XX 03-FEB-1997; 97US-0792019.  
 PF  
 XX 03-FEB-1997; 97US-0792019.  
 PR  
 XX (AMGE-) AMGEN INC.  
 PA  
 XX Chang M;  
 PI  
 XX WPI; 1998-260526/23.  
 DR N-PSDB; AAV22654.  
 DR  
 XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -  
 PT useful for stimulating growth of motor and sympathetic neurons  
 XX  
 XX Claim 2; Fig 5; 41pp; English.  
 XX  
 XX The present sequence represents a murine neurotrophic factor, designated  
 CC NNT-1, which is capable of stimulating growth of motor or sympathetic  
 CC neurons. The NNT-1 protein is useful in the treatment of neurological  
 CC diseases characterised by the degeneration and death of particular  
 CC classes of neurons. These diseases specifically include Parkinson's  
 CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,  
 CC stroke and various degenerative disorders affecting vision.  
 XX  
 SQ Sequence 225 AA;

Query Match 100.0%; Score 1224; DB 19; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-121;  
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGMACLTVMHLPAVPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60  
 Db 1 MDLRAGDSWGMACLTVMHLPAVPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60  
 QY 61 YLYNLGPPNEPDPNPRLGAEITLPRATVNLVWRSINDRLRLTONYEAYSHLLCYLRGL 120  
 Db 61 YLYNLGPPNEPDPNPRLGAEITLPRATVNLVWRSINDRLRLTONYEAYSHLLCYLRGL 120  
 QY 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180  
 Db 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180  
 QY 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPPAASVTILHLEAHGF 225  
 Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPPAASVTILHLEAHGF 225

RESULT 3  
 AAY87814  
 ID AAY87814 standard; Protein; 225 AA.  
 XX  
 AC AAY87814;  
 XX

DT 24-AUG-2000 (first entry)  
 XX  
 XX Murine NNT-1 protein.  
 DE

XX NNT-1; neurotrophic factor; nootropic; neuroprotective; treatment;  
 KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;  
 KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; murine;  
 KW Huntington's disease; peripheral neuropathy; neural retina degeneration;



KW	retinopathy; immune disorder; hematopoietic disorder.	AC	AAU78177;	
XX		XX	05-JUN-2002 (first entry)	
OS	Mus sp.	XX	Mouse novel neurotrophic factor NNT1.	
PN	US6054294-A.	DE		
XX		XX	Mouse; NNT1; neurotrophic factor; Ige-related disease; Type I allergic disease; allergic rhinitis; eczema; dermatitis; pollinosis; asthma; immune disease; cancer; arteriosclerosis; vascular restenosis; rheumatoid arthritis; psoriatic arthritis; inflammatory arthritis; osteoarthritis; inflammatory joint disease; autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis; inflammatory bowel disease; transplant rejection; reproductive disorder; graft versus host disease; infertility; miscarriage; preterm labour.	
PD	25-APR-2000.	XX	Mus sp.	
XX		XX	WO200215977-A2.	
PF	12-DEC-1997; 97US-0988919.	XX	28-FEB-2002.	
XX		XX	17-AUG-2001; 2001WO-US25906.	
PR	03-FEB-1997; 97US-0792019.	XX	18-AUG-2000; 2000US-226436P.	
XX		XX	16-AUG-2001; 2001US-0931704.	
PA	(AMGE-) AMGEN INC.	XX	(AMGE-) AMGEN INC.	
XX		XX	Senaldi G;	
PI	Chang M;	XX	WPI; 2002-280967/32.	
DR	WPI; 2000-338492/29.	XX	N-PSDB; ABK11649.	
DR	N-PSDB; AAA39483.	XX	Treating Immunoglobulin E-related disease, modulating Ige levels in a patient, preventing Ige-related disease and treating allergic diseases, involves administering NNT-1 inhibitor to a patient	
XX	New nucleic acids encoding neurotrophic factors useful for stimulating growth of motor or sympathetic neurons for treating neuron cell damage	XX	Claim 2; Fig 5; 63pp; English.	
PT		XX	The invention relates to treating Immunoglobulin E (Ige)-related disease, modulating Ige levels in a patient, preventing an Ige-related disease, and treating allergic diseases, comprising administering a therapeutically effective amount of novel neurotrophic factor (NNT)-1 inhibitor to a patient. Also included are a method of diagnosing an Ige-related disease or susceptibility to an Ige-related disease, by determining the presence or amount of expression of an NNT1 polypeptide encoded by a NNT1 nucleotide sequence, its fragment or naturally occurring variant, and diagnosing an Ige-related disease or susceptibility of an Ige-related disease based on the presence or amount of expression of the polypeptide and a pharmaceutical composition for use in treating Ige-related disease, comprising the NNT1 inhibitor.	
PS	Claim 2b; Fig 5; 42pp; English.	XX	The NNT1 inhibitor is useful for preventing and treating Ige-related disease, modulating Ige levels, and treating allergic diseases e.g. Type I allergic disease, allergic rhinitis, eczema, dermatitis, pollinosis, asthma, immune diseases and disorders, diseases involving abnormal cell proliferation including cancer, arteriosclerosis and vascular restenosis, diseases and conditions relating to dysfunction of immune system including rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, osteoarthritis, inflammatory joint disease, autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory bowel disease, transplant rejection, and graft versus host disease, and reproductive diseases and disorders including infertility, miscarriage, preterm labour and delivery, and endometriosis. The present sequence represents Mouse NNT1.	
XX		XX	Sequence 225 AA;	
XX		XX	Query Match 100.0%; Score 1224; DB 23; Length 225;	
XX		XX	Best Local Similarity 100.0%; Pred. No. 2.4e-121; Mismatches 0; Indels 0; Gaps 0;	
XX		XX	Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
XX		XX	1 MDLRAGDSWGLACLTCLVTLVHLPVAPALNRTGDPGPGPSIQKTYDLYLEHQLSLAGT 60	
XX		XX	1 MDLRAGDSWGLACLTCLVTLVHLPVAPALNRTGDPGPGPSIQKTYDLYLEHQLSLAGT 60	
XX		XX	61 YLYNLGPPFNEPDPFPRGAGTTLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120	
XX		XX	61 YLYNLGPPFNEPDPFPRGAGTTLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120	
XX		XX	121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTATLGYPLPQPLPGTEPAWAPGPAHSDFIQ 180	
XX		XX	121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTATLGYPLPQPLPGTEPAWAPGPAHSDFIQ 180	
XX		XX	181 KMDDFWLLKELOTWLRSAKDNRLKKKQPPAASVTLLHLEAHGF 225	
XX		XX	181 KMDDFWLLKELOTWLRSAKDNRLKKKQPPAASVTLLHLEAHGF 225	
XX		XX	RESULT 4	
XX		XX	AAU78177	
XX		XX	ID AAU78177 standard; Protein; 225 AA.	

Db 1 MDLRAGDSWGMGLACLTVMHLPAVPALNRTGDPGPGPSIQKTYDLTRYLHQLRSLAGT 60  
 QY 61 YLYNLGPPPEPNPRLGAEITLPRATVNLVWRSNDRLRLTONYAYSHLLCYLRGL 120  
 Db 61 YLYNLGPPPEPNPRLGAEITLPRATVNLVWRSNDRLRLTONYAYSHLLCYLRGL 120  
 QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQPLPGTEPAWPGPAHSDFLQ 180  
 Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQPLPGTEPAWPGPAHSDFLQ 180  
 QY 181 KMDDFWLLKELOTLWRSKDFNRLKKMQPPAAASVTLHLEAHGF 225  
 Db 181 KMDDFWLLKELOTLWRSKDFNRLKKMQPPAAASVTLHLEAHGF 225

## RESULT 5

AAW29715  
 ID AAW29715 standard; Protein; 225 AA.

AC AAW29715;

DT 09-NOV-1998 (first entry)

DE Human neurotrophic factor NNT-1.

XX NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
 KW peripheral neuropathy; dystrophy; neural retina degeneration;  
 KW common variable immunodeficiency; CVID; selective IGA deficiency;  
 KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;  
 KW therapy.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Peptide 1..27

FT Protein /label= Sig\_peptide

FT /label= Mat\_protein

XX WO9833922-A1.

XX 06-AUG-1998.

PD 02-FEB-1998; 98WO-US02363.

PF 30-JAN-1998; 98US-0016534.

PR 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.

XX Chang M, Elliot GS, Sarmiento U, Senaldi G;

XX WPI; 1998-437475/37.

XX N-PSDB; AAV47510-11.

XX Newly isolated nucleic acid encoding human or murine neurotrophic

XX factor NNT-1 - useful for treatment of neurological and

XX immunological diseases or inflammation, also as vaccine adjuvant

XX Claim 12; Fig 3; 120pp; English.

XX This is the amino acid sequence of a novel neurotrophic factor,

XX designated NNT-1, that is a growth factor for neurons and for B or

XX T cells. It was deduced from isolated cDNA (see AAV47510) and

XX genomic DNA (see AAV47511) clones. Vectors containing the cDNA or

XX genomic DNA and host cells are provided for use in the production

XX of NNT-1 polypeptides. These are used to treat: (i) neurological

XX or immunological diseases, specifically Alzheimer's, Parkinson's

XX or Huntington's diseases, amyotrophic lateral sclerosis,

XX Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and

XX degeneration of the neural retina, or conditions characterised by T

CC or B cell defects, e.g. common variable immunodeficiency (CVID),  
 CC selective IGA deficiency, hypogammaglobulinaemia and X-linked  
 CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)  
 CC inflammation. NNT-1 is also able to boost immunoreactivity and  
 CC antibody production following vaccination, and, since it inhibits  
 CC tumour necrosis factor production, it may also be useful for  
 CC treating sepsis. In addition, cells that have been engineered to  
 CC express NNT-1 can be implanted, or nucleic acids are delivered in  
 CC gene therapy vectors.

XX Sequence 225 AA;

Query Match 97.5%; Score 1193; DB 19; Length 225;

Best Local Similarity 96.9%; Pred. No. 4.7e-118;

Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGMGLACLTVMHLPAVPALNRTGDPGPGPSIQKTYDLTRYLHQLRSLAGT 60

Db 1 MDLRAGDSWGMGLACLTVMHLPAVPALNRTGDPGPGPSIQKTYDLTRYLHQLRSLAGT 60

QY 61 YLYNLGPPPEPNPRLGAEITLPRATVNLVWRSNDRLRLTONYAYSHLLCYLRGL 120

Db 61 YLYNLGPPPEPNPRLGAEITLPRATVNLVWRSNDRLRLTONYAYSHLLCYLRGL 120

QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQPLPGTEPAWPGPAHSDFLQ 180

Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQPLPGTEPAWPGPAHSDFLQ 180

QY 181 KMDDFWLLKELOTLWRSKDFNRLKKMQPPAAASVTLHLEAHGF 225

Db 181 KMDDFWLLKELOTLWRSKDFNRLKKMQPPAAASVTLHLEAHGF 225

## RESULT 6

AAW56141

ID AAW56141 standard; Protein; 225 AA.

XX AAW56141;

XX 13-JUL-1998 (first entry)

XX Amino acid sequence of human neurotrophic factor NNT-1.

XX Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;

XX treatment; neurological disease; degeneration; Parkinson's disease;

XX amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..27

XX Protein /note= "signal peptide"

XX /note= "mature protein"

XX US5741772-A.

XX 21-APR-1998.

XX 03-FEB-1997; 97US-0792019.

XX 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.

XX Chang M;

XX WPI; 1998-260526/23.

XX N-PSDB; AAV22652.

XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -

XX useful for stimulating growth of motor and sympathetic neurons

PS Claim 1; Fig 3; 4lpp; English.

XX The present sequence represents a human neurotrophic factor, designated  
CC NNT-1, which is capable of stimulating growth of motor or sympathetic  
CC neurons. The NNT-1 protein is useful in the treatment of neurological  
CC diseases characterized by the degeneration and death of particular  
CC classes of neurons. These diseases specifically include Parkinson's  
CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,  
CC stroke and various degenerative disorders affecting vision.

XX SQ Sequence 225 AA;

Query Match 97.5%; Score 1193; DB 19; Length 225;

Best Local Similarity 96.9%; Pred. No. 4.7e-118;

Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACLTCTVWLHLPVAPALNRTGDPGPGPSIQKTYDLTRYLHQLSLAGT 60

Db 1 MDLRAGDSWGLACLTCTVWLHLPVAPALNRTGDPGPGPSIQKTYDLTRYLHQLSLAGT 60

Qy 61 YLYLGGPPFNPDPNPPRLGAEETLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120

Db 61 YLYLGGPPFNPDPNPPRLGAEETLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120

Qy 121 NRQAATAEILRRSLAHFCTSLQGLGSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180

Db 121 NRQAATAEILRRSLAHFCTSLQGLGSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLEAHGF 225

Db 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLEAHGF 225

RESULT 7

AAW94466

ID AAW94466 standard; Protein; 225 AA.

XX AC AAW94466;

DT 22-APR-1999 (first entry)

XX Human cardiotrophin-like cytokine protein.

XX Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;  
KW CLC; IL-6; diagnosis; detection; immune system-related disorder;  
KW cancer; cardiac disorder; heart failure; hypertension; cancer;  
KW autoimmune disorder; infection.

XX OS Homo sapiens.

XX FH Location/Qualifiers

FT Peptide 1..27

FT Protein /label= signal

FT Domain /label= Cardiotrophin-like\_cytokine

FT Domain /label= CD-I

FT Domain /note= "conserved domain"

FT Domain /label= CD-II

FT Domain /note= "conserved domain"

FT Domain /label= CD-III

FT Domain /note= "conserved domain"

XX PN WO9900415-A1.

XX PD 07-JAN-1999.

XX PF 29-JUN-1998; 98WO-US13129.

XX PR 30-JUN-1997; 97US-0051311.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Shi Y;

XX WPI; 1999-095678/08.

DR N-PSDB; AAX16161.

XX New isolated cardiotrophin-like cytokine nucleic acid - used to

PT develop products for treating cardiac and immune system disorders,

PT e.g. heart failure, hypertension, cancers, autoimmune disorders and

PT infections

XX Claim 1; Fig 1; 103pp; English.

XX The present invention relates to a novel cardiotrophin-like cytokine

XX (CLC) protein which is a member of the interleukin 6 (IL-6) cytokine

XX family. The present sequence represents the human CLC protein. The

XX present invention also describes screening methods for identifying

XX agonists and antagonists of CLC activity, as well as methods for

XX detecting cardiac and immune system-related disorders and

XX therapeutic methods for treating cardiac and immune system-related

XX disorders, e.g. heart failure, hypertension, cancers, autoimmune

XX disorders and infections.

XX SQ Sequence 225 AA;

Query Match 97.5%; Score 1193; DB 20; Length 225;

Best Local Similarity 96.9%; Pred. No. 4.7e-118;

Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACLTCTVWLHLPVAPALNRTGDPGPGPSIQKTYDLTRYLHQLSLAGT 60

Db 1 MDLRAGDSWGLACLTCTVWLHLPVAPALNRTGDPGPGPSIQKTYDLTRYLHQLSLAGT 60

Qy 61 YLYLGGPPFNPDPNPPRLGAEETLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120

Db 61 YLYLGGPPFNPDPNPPRLGAEETLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120

Qy 121 NRQAATAEILRRSLAHFCTSLQGLGSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180

Db 121 NRQAATAEILRRSLAHFCTSLQGLGSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLEAHGF 225

Db 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLEAHGF 225

RESULT 8

AAW87813

ID AAW87813 standard; Protein; 225 AA.

XX AC AAW87813;

XX DT 24-AUG-2000 (first entry)

XX DE Human NNT-1 protein.

XX KW NNT-1; human; neurotrophic factor; neurotropic; neuroprotective; treatment;

KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;

KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;

KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;

KW Huntington's disease; peripheral neuropathy; neural retina degeneration;

KW retinopathy; immune disorder; hematopoietic disorder.

XX OS Homo sapiens.

XX PN US6054294-A.

XX PD 25-APR-2000.

XX PF 12-DEC-1997; 97US-0988819.

XX PR 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.  
 PA Chang M;  
 PI WPI; 2000-338492/29.  
 XX N-PSDB; AAA39481.  
 DR New nucleic acids encoding neurotrophic factors useful for stimulating  
 PT growth of motor or sympathetic neurons for treating neuron cell damage  
 PT  
 XX Claim 1c; Fig 3; 42pp; English.  
 XX This invention describes a novel nucleic acid molecule (I) encoding a  
 CC novel neurotrophic factor (NNT-1) (II) which has neurotrophic, neuroprotective,  
 CC anticonvulsant, antiparkinsonian, antidiabetic and ophthalmological activity. (I) is useful for producing NNT-1  
 CC polypeptides which are useful for treating patients in whom various  
 CC cells of the central, autonomic, or peripheral nervous system have  
 CC degenerated and/or have been damaged by congenital disease, trauma,  
 CC mechanical damage, surgery, stroke, ischemia, infection, metabolic  
 CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1  
 CC proteins are used to treat diseases like Alzheimer's, Parkinson's,  
 CC amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's  
 CC disease, peripheral neuropathy induced by diabetes or other metabolic  
 CC disorders, and/or dystrophies or degeneration of the neural retina such  
 CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of  
 CC night blindness, progressive cone-rod degeneration, immune disorders and  
 CC hematopoietic disorders. (II) is effective in treating neurological  
 CC conditions and promotes neuron regeneration. Neural functions are  
 CC effectively restored in patients suffering from various neurological  
 CC disorders. This sequence represents the human NNT-1 protein described in  
 CC the method of the invention.  
 XX  
 SQ Sequence 225 AA;  
 Query Match 97.5%; Score 1193; DB 21; Length 225;  
 Best Local Similarity 96.9%; Pred. No. 4.7e-118;  
 Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MDLRAGDSWGMACLCCTVLMHLPVAPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGT 60  
 Db 1 MDLRAGDSWGMACLCCTVLMHLPVAPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGT 60  
 QY 61 YLNYLGPPFPNEPDPNPPRLGAETLPVATVLEWRSNDRLRLTONYEAISHLLCYLRGL 120  
 Db 61 YLNYLGPPFPNEPDPNPPRLGAETLPVATVLEWRSNDRLRLTONYEAISHLLCYLRGL 120  
 QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTGLGYPLPQPLPGTEPAPAGPAHSDFLQ 180  
 Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTGLGYPLPQPLPGTEPAPAGPAHSDFLQ 180  
 QY 181 KMDDFWLLKELQTLWRSKDFNRLKKMKOPPAASVTLHLEAHGF 225  
 Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKOPPAASVTLHLEAHGF 225  
 RESULT 9  
 AAG63543  
 ID AAG63543 standard; Protein; 225 AA.  
 XX  
 AC AAG63543;  
 XX  
 DT 15-OCT-2001 (first entry)  
 XX  
 DE Amino acid sequence of a human NNT-1 protein.  
 XX  
 KW NNT-1; CLF-1; sCNTFRalpha; nervous system; neuron; nervous system;  
 KW neuro-muscular function; tumour; immune system; haematopoietic system;  
 KW reproductive system; liver; skeletal muscle; neurodegenerative disease;  
 KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;  
 KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;

KW blastocyst implantation; thrombosis; retinal disease;  
 KW retinal pigmentosis.  
 XX Homo sapiens.  
 XX WO200155172-A2.  
 XX 02-AUG-2001.  
 XX 26-JAN-2001; 2001WO-FR00253.  
 XX 27-JAN-2000; 2000FR-0001035.  
 XX 12-OCT-2000; 2000FR-0013089.  
 XX (FABR ) FABRE MEDICAMENT SA PIERRE.  
 XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;  
 XX WPI; 2001-488773/53.  
 XX N-PSDB; AAH74484.  
 XX A complex comprising a NNT-1 protein and a CLF-1 and/or sCNTFRalpha  
 PT protein useful to treat neurodegenerative disease including Parkinson's  
 PT and Huntington's, obesity and cancer -  
 XX Claim 2; Page 58; 67pp; French.  
 XX The present sequence represents a human NNT-1 protein. The specification  
 CC describes a complex comprising a NNT-1 protein and a CLF-1 and/or  
 CC sCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate  
 CC activity of the sCNTFRalpha/gp130/LIFRbeta receptor complex, or to  
 CC induce phosphorylation of the tyrosine of gp130 and LIFRbeta,  
 CC particularly where cells expressing the receptor complex are in the  
 CC central or peripheral nervous system, in neurons implicated in  
 CC neuro-muscular function or in skeletal muscle. The complex or  
 CC antibodies are also used to decrease the survival, growth or  
 CC proliferation of tumour cells or to facilitate the proliferation and/or  
 CC inhibit differentiation of cells stocks. The complex is also used to  
 CC modulate activity of the gp130/LIFRbeta receptor or cells expressing  
 CC that receptor, particularly those cells implicated in the immune,  
 CC haematopoietic, nervous or reproductive system, the liver or skeletal  
 CC muscle. Molecules of the invention may be used to prevent or treat  
 CC neurodegenerative diseases including amyotrophic lateral sclerosis,  
 CC Parkinson's and Huntington's disease, to repair or regenerate nervous  
 CC or muscular tissue or to maintain muscular mass in paralysis patients.  
 CC They may also be used to treat cancer, obesity and associated diseases,  
 CC and to improve fertility, particularly to avoid endometriosis and/or  
 CC assist blastocyst implantation, thrombosis, or retinal disease,  
 CC particular retinal pigmentosis.  
 XX  
 SQ Sequence 225 AA;  
 Query Match 97.5%; Score 1193; DB 22; Length 225;  
 Best Local Similarity 96.9%; Pred. No. 4.7e-118;  
 Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MDLRAGDSWGMACLCCTVLMHLPVAPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGT 60  
 Db 1 MDLRAGDSWGMACLCCTVLMHLPVAPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGT 60  
 QY 61 YLNYLGPPFPNEPDPNPPRLGAETLPVATVLEWRSNDRLRLTONYEAISHLLCYLRGL 120  
 Db 61 YLNYLGPPFPNEPDPNPPRLGAETLPVATVLEWRSNDRLRLTONYEAISHLLCYLRGL 120  
 QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTGLGYPLPQPLPGTEPAPAGPAHSDFLQ 180  
 Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTGLGYPLPQPLPGTEPAPAGPAHSDFLQ 180  
 QY 181 KMDDFWLLKELQTLWRSKDFNRLKKMKOPPAASVTLHLEAHGF 225  
 Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKOPPAASVTLHLEAHGF 225

RESULT 10  
AAU78176  
ID AAU78176 standard; Protein; 225 AA.  
XX AC  
XX AAU78176;  
XX 05-JUN-2002 (first entry)  
XX DE Human novel neurotrophic factor NNT1.  
XX KW Human; NNT1; neurotrophic factor; IGE-related disease;  
KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;  
KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;  
KW vascular restenosis; rheumatoid arthritis; psoriatic arthritis;  
KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;  
KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;  
KW inflammatory bowel disease; transplant rejection; reproductive disorder;  
KW graft versus host disease; infertility; miscarriage; preterm labour.  
XX OS Homo sapiens.  
XX KW  
XX PN W0200215977-A2.  
XX PD  
XX PD 28-FEB-2002.  
XX PF 17-AUG-2001; 2001WO-US25906.  
XX PR 18-AUG-2000; 2000US-226436P.  
XX PR 16-AUG-2001; 2001US-0931704.  
XX KW  
XX PA (AMGE-) AMGEN INC.  
XX PI Senaldi G;  
XX DR WPI; 2002-280967/32.  
XX DR N-PSDB; ABK11647.  
XX PT Treating Immunoglobulin E-related disease, modulating IGE levels in a  
PT patient, preventing IGE-related disease and treating allergic diseases,  
PT involves administering NNT-1 inhibitor to a patient -  
XX Claim 2; Fig 3; 63pp; English.  
XX CC The invention relates to treating Immunoglobulin E (IGE)-related disease,  
CC modulating IGE levels in a patient, preventing an IGE-related disease,  
CC and treating allergic diseases, comprising administering a  
CC therapeutically effective amount of novel neurotrophic factor (NNT)-1  
CC inhibitor to a patient. Also included are a method of diagnosing an  
CC IGE-related disease or susceptibility to an IGE-related disease, by  
CC determining the presence or amount of expression of an NNT1 polypeptide  
CC encoded by a NNT1 nucleotide sequence, its fragment or naturally  
CC occurring variant, and diagnosing an IGE-related disease or  
CC susceptibility of an IGE-related disease based on the presence or amount  
CC of expression of the polypeptide and a pharmaceutical composition for use  
CC in treating IGE-related disease, comprising the NNT1 inhibitor.  
CC The NNT1 inhibitor is useful for preventing and treating IGE-related  
CC disease, modulating IGE levels, and treating allergic diseases e.g.  
CC Type I allergic disease, allergic rhinitis, eczema, dermatitis,  
CC pollinosis, asthma, immune diseases and disorders, diseases involving  
CC abnormal cell proliferation including cancer, arteriosclerosis and  
CC vascular restenosis, diseases and conditions relating to dysfunction of  
CC immune system including rheumatoid arthritis, psoriatic arthritis,  
CC inflammatory arthritis, osteoarthritis, inflammatory joint disease,  
CC autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory  
CC bowel disease, transplant rejection, and graft versus host disease, and  
CC reproductive diseases and disorders including infertility, miscarriage,  
CC preterm labour and delivery, and endometriosis. The present sequence  
CC represents human NNT1.  
XX SQ Sequence 225 AA;  
Query Match 97.5%; Score 1193; DB 23; Length 225;

Best Local Similarity 96.9%; Pred. No. 4.7e-118;  
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDLRAGDSGMLACICTVLWHLPAVPALNRTGDPGPSIOKTYDLTRYLHQLRSLAGT 60  
DB 1 MDLRAGDSGMLACICTVLWHLPAVPALNRTGDPGPSIOKTYDLTRYLHQLRSLAGT 60  
QY 61 YLNYLGPPFNEPDNPPRLGAETLPRATVNLVWRSNDLRLTQNYEAYSHLLCYLRGL 120  
DB 61 YLNYLGPPFNEPDNPPRLGAETLPRATVNLVWRSNDLRLTQNYEAYSHLLCYLRGL 120  
QY 121 NROQATAELRRSLAHFCTSLQGLLSGSIAGVWATLGYPLPQLPCTEPAWAPGPAHSDFLQ 180  
DB 121 NROQATAELRRSLAHFCTSLQGLLSGSIAGVWATLGYPLPQLPCTEPAWAPGPAHSDFLQ 180  
QY 181 KMDDFWLLKELQWLWRSADFNRLKKMOPPAASVTLLHLEAHGF 225  
DB 181 KMDDFWLLKELQWLWRSADFNRLKKMOPPAASVTLLHLEAHGF 225

## RESULT 11

AAU25831  
ID AAU25831 standard; Protein; 253 AA.

XX AC AAU25831;

XX DT 16-OCT-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:1346.

XX KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiac; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;  
KW dermatological; anti-allergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW anaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder.

XX OS Homo sapiens.

XX PN W0200153455-A2.

XX PD 26-JUL-2001.

XX PF 22-DEC-2000; 2000WO-US35017.

XX PR 23-DEC-1999; 99US-0471275.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-457603/49.

XX DR N-PSDB; AAH99772.

XX PT Isolated human polynucleotides encoding polypeptides, useful for the  
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX Claim 20; Page 278; 1217pp; English.  
XX PS  
XX CC AAH99166 to AAH99904 encode the human proteins given in AAU25225 to  
XX AAU25963. The proteins can have activities based on the tissues and  
XX cells they are expressed in, such as: antiinflammatory; antirheumatic;

antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; viricide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmunity, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders.

XX Sequence 253 AA;

Query Match 97.5%; Score 1193; DB 22; Length 253;  
Best Local Similarity 96.9%; Pred. No. 5.5e-118;  
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMACLCVTLVHLPVAVPALNRTGDPGPGPSIQKTYDTRYLHQLRSLAGT 60  
Db 29 MDLRAGDSWGMACLCVTLVHLPVAVPALNRTGDPGPGPSIQKTYDTRYLHQLRSLAGT 88  
Qy 61 YLYNLGPPNEPDNPRLGAEITLPATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120  
Db 89 YLYNLGPPNEPDNPRLGAEITLPATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 148  
Qy 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQLPGTEPAWAPGPAHSDFLQ 180  
Db 149 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQLPGTEPAWAPGPAHSDFLQ 208  
Qy 181 KMDDFWLLKELOTWLRSAKDFNRLKKMQPPAAAVTLHLEAHGF 225  
Db 209 KMDDFWLLKELOTWLRSAKDFNRLKKMQPPAAAVTLHLEAHGF 253

RESULT 12

AAE00828  
ID AAE00828 standard; Protein; 223 AA.

XX AAE00828;

XX 02-JUL-2001 (first entry)

XX Human cardiotrophin-like cytokine (CLC) protein.

XX Human; biologically active complex; haemopoietin receptor; NR6;  
XX cardiotrophin-like cytokine; CLC; therapy; prophylaxis; proliferation;  
XX differentiation; cell survival; neurotrophic activity.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..27

XX /label= Signal\_peptide

XX Protein 28..223

XX /label= Human\_mature\_CLC\_protein

XX /note= "Cardiotrophin-like cytokine"

XX WO200127157-A1.

XX 19-APR-2001.

XX 06-OCT-2000; 2000WO-AU01216.

XX 08-OCT-1999; 99AU-0003327.

XX 12-MAY-2000; 2000AU-0007489.

XX

PA (AMRA-) AMRAD OPERATIONS PTY LTD.

XX Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;

PI Nakata Y, Hasegawa M;

XX WPI; 2001-281978/29.

DR N-PSDB; AAD04201.

XX New biologically active complex comprising NR6 and

PT cardiotrophin-like cytokine, for facilitating proliferation,

PT differentiation and/or survival of a cell -

XX Claim 32; Page 114-115; 123pp; English.

XX The present invention relates to a biologically active complex comprising  
CC a haemopoietin receptor, NR6 and cardiotrophin-like cytokine (CLC).  
CC The complex is useful in the manufacture of a medicament for the  
CC treatment and/or prophylaxis of a subject, as it is involved in  
CC facilitating proliferation, differentiation and/or survival of a cell.  
CC The complex or its components have neurotrophic activity. The present  
CC sequence is human cardiotrophin-like cytokine (CLC) protein.

XX Sequence 223 AA;

Query Match 96.5%; Score 1181; DB 22; Length 223;  
Best Local Similarity 96.9%; Pred. No. 8.7e-117;  
Matches 216; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMACLCVTLVHLPVAVPALNRTGDPGPGPSIQKTYDTRYLHQLRSLAGT 60

Db 1 MDLRAGDSWGMACLCVTLVHLPVAVPALNRTGDPGPGPSIQKTYDTRYLHQLRSLAGT 60

Qy 61 YLYNLGPPNEPDNPRLGAEITLPATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120

Db 61 YLYNLGPPNEPDNPRLGAEITLPATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120

Qy 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQLPGTEPAWAPGPAHSDFLQ 180

Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQLPGTEPAWAPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELOTWLRSAKDFNRLKKMQPPAAAVTLHLEAH 223

Db 181 KMDDFWLLKELOTWLRSAKDFNRLKKMQPPAAAVTLHLEAH 223

RESULT 13

ABE11896

XX ABE11896 standard; peptide; 321 AA.

XX ABE11896;

XX 11-JAN-2002 (first entry)

XX Human cardiotrophin-like cytokine homologue, SEQ ID NO:2266.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
XX haematopoiesis regulation; tissue growth; immunomodulator; activin;  
XX inhibin; chemotaxis; chemokines; thrombolysis; oncogenesis;  
XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
XX myeloid cell disorder; lymphoid cell disorder; athma; arthritis;  
XX chronic inflammatory condition; proliferative retinopathy;  
XX atherosclerosis; coronary heart disease; arterial ischaemia;  
XX bone disorder; osteoporosis; vascular growth disorder;  
XX tissue regeneration; wound healing; infection; immune disorder;  
XX cell culture; drug screening; gene therapy; antiinflammatory;  
XX antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
XX cytostatic; osteopathic; vasotropic; cardiac; virucide; antibacterial;  
XX antifungal; vulnerary; antiulcer.

OS Homo sapiens.

XX WO200157188-A2.

PN

XX PD 09-AUG-2001.  
 XX PF 05-FEB-2001; 2001WO-US03800.  
 XX PR 03-FEB-2000; 2000US-0496914.  
 XX PR 27-APR-2000; 2000US-0560875.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Drmanac RT;  
 XX DR WPI; 2001-457740/49.  
 XX DR N-PSDB; ABA09140.  
 XX PT Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX PS Claim 20; Page 273; 1963pp; English.  
 XX CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.

XX Sequence 321 AA;

Query Match 95.7%; Score 1171; DB 22; Length 321;  
 Best Local Similarity 96.0%; Pred. No. 1.6e-115;  
 Matches 214; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 LRAGDSWGMACLCVTLVHLPVAPALNRTGDPGPGPSIQKTVLDLTYLEHQLRSLAGTYL 62

Db 99 LPTGDSWGMACLCVTLVHLPVAPALNRTGDPGPGPSIQKTVLDLTYLEHQLRSLAGTYL 158

QY 63 NYLGPPFPNPPRLGHTLPVATVNLVWRSNDRLRLTQVYAYSHLLCYLRLGNR 122

Db 159 NYLGPPFPNPPRLGHTLPVATVNLVWRSNDRLRLTQVYAYSHLLCYLRLGNR 218

QY 123 QAATAELRRSLAHFCTSLQGLLSIAGVMATLGYPLPQPLPCTEPANAPGPAHSDFLQKM 182  
 Db 219 QAATAELRRSLAHFCTSLQGLLSIAGVMATLGYPLPQPLPCTEPANAPGPAHSDFLQKM 278  
 QY 183 DDFWLLKELQTLWRSKADFNRLKKKQPPAASVTLHLEAHGF 225  
 Db 279 DDFWLLKELQTLWRSKADFNRLKKKQPPAASVTLHLEAHGF 321

# RESULT 14

AAW79399  
 ID AAW79399 standard; Protein; 321 AA.  
 XX AC AAW79399;  
 XX DT 06-NOV-2001 (first entry)  
 XX DE Human protein SEQ ID NO 3045.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

XX N-PSDB; AAK52532.

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
 useful in diagnosis and gene therapy -

XX Claim 20; Page 237; 621pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.

XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAW80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.

XX Sequence 321 AA;

Query Match 95.7%; Score 1171; DB 22; Length 321;

Best Local Similarity 96.0%; Pred. No. 1.6e-115;  
Matches 214; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 LRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLRSLAGTYL 62  
DB 99 LFTGDSWGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLRSLAGTYL 158

QY 63 NYLGPPFNEPDPNPPRLGAETLPRATVNLVWRSNDRLRLTQNYEAYSHLLCYLRGLNR 122  
DB 159 NYLGPPFNEPDPNPPRLGAETLPRATVNLVWRSNDRLRLTQNYEAYSHLLCYLRGLNR 218

QY 123 QAATAELRSLAHFCTSLQGLLSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLOKM 182  
DB 219 QAATAELRSLAHFCTSLQGLLSIAGVMATLGYPLPQPLPGTEPTWTPGPAHSDFLOKM 278

QY 183 DDFWLLKELQTLWRSKDFNRLKKMQPPAAASVTLHLEAHGF 225  
DB 279 DDFWLLKELQTLWRSKDFNRLKKMQPPAAAVTLHLEAHGF 321

RESULT 15  
AAB19587  
ID AAB19587 standard; Protein; 215 AA.  
AC AAB19587;  
XX  
DT 22-JAN-2001 (first entry)  
XX  
DE Mouse interleukin-B60 (IL-B60).  
XX  
KW Interleukin-B60; IL-B60; mouse; cytokine; cytokine-like factor-1;  
KW haematopoietic; inflammation; antiinflammatory; autoimmune disease;  
KW therapy.  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..17  
FT /label= Signal\_peptide  
FT Protein 18..215  
FT /label= Mature-protein  
XX  
PN WO200053631-A1.  
XX  
PD 14-SEP-2000.  
XX  
PF 09-MAR-2000; 2000WO-US06182.  
XX  
PR 11-MAR-1999; 99US-0267901.  
XX  
FA (SCHE ) SCHERING CORP.  
XX  
PI Oppmann B, Timans JC, Kastelein RA, Bazan JF;  
XX  
DR WPI; 2000-587426/55.  
DR N-PSDB; AAA88547.  
XX  
FT Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes,  
FT polypeptides, and nucleic acids, useful in research, diagnosis and for  
FT treating inflammatory and autoimmune disorders -  
XX  
PS Claim 1; Page 17; 97pp; English.  
XX  
CC The present sequence is that of mouse interleukin-B60 (IL-B60), a  
CC novel, small soluble cytokine-like protein that exhibits structural  
CC motifs characteristic of a member of the long-chain cytokines, and  
CC which shows homology to granulocyte colony stimulating factor and  
CC interleukin-6. IL-608 may have either stimulatory or inhibitory  
CC effects on haematopoietic cells, including e.g. lymphoid cells,  
CC such as T-cells, B-cells, natural killer cells, macrophages,  
CC dendritic cells, haematopoietic progenitors, etc. Methods are  
CC provided for modulating the physiology or development of a cell or  
CC tissue culture cells by contacting the cell with an agonist or

CC antagonist of IL-B60 or an agonist of antagonist of a complex of  
CC mature IL-B60 and its partner, cytokine-like factor-1 (CLF-1, see  
CC AAB19588). The IL-B60/CLF-1 cytokine serves as a key physiological  
CC factor in motor neuron development and regeneration. IL-608, its  
CC agonists and antagonists may be used to treat inflammatory or  
CC autoimmune disorders and also for drug screening.  
XX  
SQ Sequence 215 AA;  
Query Match 95.3%; Score 1167; DB 21; Length 215;  
Best Local Similarity 100.0%; Pred. No. 2.5e-115;  
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 MLACLCTVWLHLPVAPALNRTGDPGPGPSIQKTYDLTRYLEHQLRSLAGTYLNYLGPPFN 70  
DB 1 MLACLCTVWLHLPVAPALNRTGDPGPGPSIQKTYDLTRYLEHQLRSLAGTYLNYLGPPFN 60

QY 71 EPDFNPPRLGAETLPRATVNLVWRSNDRLRLTQNYEAYSHLLCYLRGLNRQAATAELR 130  
DB 61 EPDFNPPRLGAETLPRATVNLVWRSNDRLRLTQNYEAYSHLLCYLRGLNRQAATAELR 120

QY 131 RSLAHFCTSLQGLLSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLOKMDDFWLLKE 190  
DB 121 RSLAHFCTSLQGLLSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLOKMDDFWLLKE 180

QY 191 LQTLWRSKDFNRLKKMQPPAAASVTLHLEAHGF 225  
DB 181 LQTLWRSKDFNRLKKMQPPAAASVTLHLEAHGF 215

Search completed: January 27, 2003, 15:41:34  
Job time : 33 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:42:42 ; Search time 9 Seconds  
(without alignments)  
504.464 Million cell updates/sec

Title: US-09-931-704-5

Perfect score: 1224

Sequence: 1 MDLRAGDSWGLACLTCLVLM.....KKKMQPPAASVTILHLEAHGF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1224	100.0	225	10	US-09-931-704-5
2	1193	97.5	225	10	US-09-931-704-2
3	852	69.6	164	10	US-09-864-761-40014
4	110.5	9.0	203	10	US-09-896-856-3
5	92.5	7.6	243	9	US-10-000-776-6
6	92.5	7.6	243	9	US-09-791-497-8
7	92.5	7.6	243	10	US-09-810-052-5
8	91.5	7.5	201	10	US-09-901-540-3
9	91.5	7.5	201	10	US-09-896-856-8
10	91.5	7.5	201	10	US-09-901-257-3
11	90	7.4	195	10	US-09-770-361-5
12	87	7.1	242	9	US-10-000-776-2
13	87	7.1	242	9	US-09-791-497-2
14	86	7.0	232	10	US-09-810-052-2
15	81.5	6.7	348	10	US-09-730-617-2
16	80.5	6.6	379	9	US-09-860-846-16
17	80.5	6.6	379	10	US-09-861-289-16
18	80.5	6.6	3782	9	US-09-860-846-4
19	80.5	6.6	3782	10	US-09-861-289-4

20	79	6.5	1620	10	US-09-827-949-2	Sequence 2, Appli
21	78.5	6.4	1399	9	US-09-388-221-4	Sequence 4, Appli
22	78.5	6.4	1424	9	US-09-388-221-12	Sequence 12, Appli
23	78.5	6.4	1429	10	US-09-996-617-2	Sequence 2, Appli
24	78.5	6.4	1429	10	US-09-931-071-2	Sequence 2, Appli
25	78.5	6.4	1443	9	US-09-388-221-6	Sequence 6, Appli
26	78.5	6.4	1454	9	US-09-388-221-10	Sequence 10, Appli
27	78.5	6.4	1473	9	US-09-388-221-2	Sequence 2, Appli
28	78	6.4	427	8	US-08-808-031A-40	Sequence 40, Appli
29	78	6.4	625	10	US-09-771-161A-242	Sequence 242, App
30	78	6.4	625	10	US-09-771-161A-243	Sequence 243, App
31	78	6.4	917	10	US-09-815-242-5603	Sequence 5603, Ap
32	78	6.4	920	10	US-09-815-242-12181	Sequence 12181, A
33	78	6.4	920	10	US-09-815-242-12995	Sequence 12995, A
34	78	6.4	920	10	US-09-815-242-13148	Sequence 13148, A
35	77.5	6.3	218	10	US-09-893-737-28	Sequence 28, Appl
36	77.5	6.3	451	10	US-09-764-864-1344	Sequence 1344, Ap
37	77	6.3	473	10	US-09-871-874-19	Sequence 19, Appl
38	76	6.2	418	9	US-09-946-807-3	Sequence 3, Appli
39	76	6.2	418	10	US-09-795-686-3	Sequence 3, Appli
40	76	6.2	418	10	US-09-795-686-3	Sequence 3, Appli
41	75.5	6.2	237	10	US-09-925-301-867	Sequence 867, App
42	75.5	6.2	334	10	US-09-953-342-24	Sequence 24, Appl
43	74.5	6.1	1009	8	US-08-987-689A-2	Sequence 2, Appli
44	74	6.0	416	9	US-10-124-429-2	Sequence 2, Appli
45	73.5	6.0	856	9	US-09-364-847-35	Sequence 35, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-931-704-5  
; Sequence 5, Application US/09931704  
; Patent No. US20020041873A1  
; GENERAL INFORMATION:  
; APPLICANT: Senaldi, Giorgio  
; TITLE OF INVENTION: Methods and Compositions for Treating Ige-Related Disease Using  
; TITLE OF INVENTION: Inhibitors  
; FILE REFERENCE: A-695  
; CURRENT APPLICATION NUMBER: US/09/931,704  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: US 60/226,436  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 225  
; TYPE: PRT  
; ORGANISM: Murine  
US-09-931-704-5

Query Match	100.0%;	Score	1224;	DB	10;	Length	225;
Best Local Similarity	100.0%;	Pred. NO.	7.6e-112;				
Matches	225;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MDLRAGDSWGLACLTCLVHLPAVPALNRGTGDPGSPSIQKTYDLYLTRYLHQLRSLAGT	60				
Db	1	MDLRAGDSWGLACLTCLVHLPAVPALNRGTGDPGSPSIQKTYDLYLTRYLHQLRSLAGT	60				
Qy	61	YLYNLGPPFPNPDNPPRLGAEITLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL	120				
Db	61	YLYNLGPPFPNPDNPPRLGAEITLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL	120				
Qy	121	NRQATAEIIRSLAHFCTSLQGLLGSITAGVWATIGYPLPQPLPGTEPAWAPGAHSDFLQ	180				
Db	121	NRQATAEIIRSLAHFCTSLQGLLGSITAGVWATIGYPLPQPLPGTEPAWAPGAHSDFLQ	180				
Qy	181	KMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAASVTILHLEAHGF	225				
Db	181	KMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAASVTILHLEAHGF	225				

;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006665	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006668	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006663	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006662	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006661	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006670	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	US 60/234,687	
;	PRIOR FILING DATE:	2000-09-21	
;	PRIOR APPLICATION NUMBER:	US 09/608,408	
;	PRIOR FILING DATE:	2000-06-30	
;	PRIOR APPLICATION NUMBER:	US 09/774,203	
;	PRIOR FILING DATE:	2001-01-29	
;	NUMBER OF SEQ ID NOS:	49117	
;	SOFTWARE:	Arnomax Sequence Listing Engine	
;	SEQ ID NO	40014	
;	LENGTH:	164	
;	TYPE:	PRT	
;	ORGANISM:	Homo sapiens	
;	FEATURE:		
;	OTHER INFORMATION:	MAP TO AC005849.1	
;	OTHER INFORMATION:	EXPRESSED IN BRAIN, S	
;	OTHER INFORMATION:	EXPRESSED IN BONE MAR	
;	OTHER INFORMATION:	EXPRESSED IN LUNG, S	
;	OTHER INFORMATION:	EXPRESSED IN ADULT L	
;	OTHER INFORMATION:	EXPRESSED IN PLACENTA	
;	OTHER INFORMATION:	EXPRESSED IN HEART, S	
;	OTHER INFORMATION:	EXPRESSED IN HELA, S	
;	OTHER INFORMATION:	EXPRESSED IN FETAL L	
;	OTHER INFORMATION:	EST_HUMAN HIT: AI75292	
;	OTHER INFORMATION:	SWISSPROT HIT: Q63080	
;	US-09-864-761-40014		
	Query Match	69.6%;	Score 852
	Best Local Similarity	95.1%;	Pred. No.
	Matches 156;	Conservative 4;	Mismatch
QY	62	LNLYGPPNEPDPFNPRLGAETLP	LNLYGPPNEPDPFNPRLGAETLP
Db	1	LNLYGPPNEPDPFNPRLGAETLP	LNLYGPPNEPDPFNPRLGAETLP
QY	122	ROATAEURLRSLAHFCTSLQGLLGS	ROATAEURLRSLAHFCTSLQGLLGS
Db	61	ROATAEURLRSLAHFCTSLQGLLGS	ROATAEURLRSLAHFCTSLQGLLGS
QY	182	MDDFWLLKELQTLWRSKDNFRLKKM	MDDFWLLKELQTLWRSKDNFRLKKM
Db	121	MDDFWLLKELQTLWRSKDNFRLKKM	MDDFWLLKELQTLWRSKDNFRLKKM
RESULT 4			
US-09-896-856-3			
;	Sequence 3,	Application US/09896856	
;	Patent No.	US20020137189A1	
;	GENERAL INFORMATION:		
;	APPLICANT:	Baker, Joffre	
;		Chien, Kenneth	
;		King, Kathleen	
;		Pennica, Diane	
;		Wood, William	
;	TITLE OF INVENTION:	Cardiac Hypertro	
;	NUMBER OF SEQUENCES:	8	
;	CORRESPONDENCE ADDRESS:		
;	ADDRESSEE:	Genentech, Inc.	
;	STREET:	460 Point San Bruno Bl	
;	CITY:	South San Francisco	
;	STATE:	California	







RESULT 12  
US-10-000-776-2  
; Sequence 2, Application US/10000776  
; Patent No. US20020164609A1  
; GENERAL INFORMATION:  
; APPLICANT: Timans, Jacqueline C.  
; APPLICANT: Pflanz, Stefan K.-H.  
; APPLICANT: Kastelein, Robert A.  
; APPLICANT: Bazan, Jose F.  
; APPLICANT: Rennick, Donna  
; APPLICANT: de Waal Malefyt, Rene  
; APPLICANT: Cheung, Jeanne  
; TITLE OF INVENTION: MAMMALIAN CYTOKINES; RELATED REAGENTS  
; FILE REFERENCE: DX01040K3  
; CURRENT APPLICATION NUMBER: US/10/000,776  
; CURRENT FILING DATE: 2001-11-30  
; PRIOR APPLICATION NUMBER: 09/791,497  
; PRIOR FILING DATE: 2001-02-22  
; PRIOR APPLICATION NUMBER: 09/627,897  
; PRIOR FILING DATE: 2000-07-27  
; PRIOR APPLICATION NUMBER: 60/146,581  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: 60/147,763  
; PRIOR FILING DATE: 1999-08-06  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 242  
; TYPE: PR1  
; ORGANISM: Homo sapiens  
US-10-000-776-2

Query Match 7.1%; Score 87; DB 9; Length 242;  
Best Local Similarity 22.6%; Pred. No. 0.36;  
Matches 55; Conservative 26; Mismatches 106; Indels 56; Gaps 8;  
  
QY 2 DLKAGDSWGMACLTCTVL-----WHLPAVPALNRTGDPGPGPSIQK-----TYDLTR 48  
DB 2 DLENNPKIGLSLLPLLLVQAGVWGFRPPG-----RPQLSLQELRREFTVSLHLAR 54  
  
QY 49 YLEHQLRSLAGTYLN-----YLGPPFNEPDNPPRLGAETLPRATVNLVWRSNDLR 100  
DB 55 KLLSEVRGQAHRAESHLPGVNLVLLP-----LG-EQLPDVSLTFQAWRLSDP 102  
  
QY 101 LRLTQNYEAYSHLLCYLRGLNRQAATAELRR-SLAHFTCSLQGLGSLAGTYMATLGYPLP 159  
DB 103 ERLCFISTTLQPPHAPLGLGTQGRWTNMRMQLWAMRLDLRLQRLRQVLAAGFNLP 162  
  
QY 160 QP-----LPGTEPAWAGPAGHSDFLQKMDDFWLLKELQTLWLRSAKDFNR 204  
DB 163 EEEEEEEEEERKGLLPALGALSGAQVSWPQLLSTYRLLSLSELVLSRAVRELL 222  
  
QY 205 LKK 207  
DB 223 LSK 225  
  
RESULT 13  
US-09-791-497-2  
; Sequence 2, Application US/09791497  
; Publication No. US20030008343A1  
; GENERAL INFORMATION:  
; APPLICANT: Timans, Jacqueline C.  
; APPLICANT: Kastelein, Robert A.  
; APPLICANT: Bazan, J. Fernando  
; APPLICANT: Pflanz, Stefan  
; TITLE OF INVENTION: Mammalian Cytokines; Related Reagents  
; FILE REFERENCE: DX01040K2  
; CURRENT APPLICATION NUMBER: US/09/791,497  
; CURRENT FILING DATE: 2001-02-22  
; PRIOR APPLICATION NUMBER: 09/627,897  
; PRIOR FILING DATE: 2000-07-27

; PRIOR APPLICATION NUMBER: 60/146,581  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: 60/147,763  
; PRIOR FILING DATE: 1999-08-06  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 242  
; TYPE: PR1  
; ORGANISM: primate; surmised Homo sapiens  
US-09-791-497-2

Query Match 7.1%; Score 87; DB 9; Length 242;  
Best Local Similarity 22.6%; Pred. No. 0.36;  
Matches 55; Conservative 26; Mismatches 106; Indels 56; Gaps 8;  
  
QY 2 DLKAGDSWGMACLTCTVL-----WHLPAVPALNRTGDPGPGPSIQK-----TYDLTR 48  
DB 2 DLENNPKIGLSLLPLLLVQAGVWGFRPPG-----RPQLSLQELRREFTVSLHLAR 54  
  
QY 49 YLEHQLRSLAGTYLN-----YLGPPFNEPDNPPRLGAETLPRATVNLVWRSNDLR 100  
DB 55 KLLSEVRGQAHRAESHLPGVNLVLLP-----LG-EQLPDVSLTFQAWRLSDP 102  
  
QY 101 LRLTQNYEAYSHLLCYLRGLNRQAATAELRR-SLAHFTCSLQGLGSLAGTYMATLGYPLP 159  
DB 103 ERLCFISTTLQPPHAPLGLGTQGRWTNMRMQLWAMRLDLRLQRLRQVLAAGFNLP 162  
  
QY 160 QP-----LPGTEPAWAGPAGHSDFLQKMDDFWLLKELQTLWLRSAKDFNR 204  
DB 163 EEEEEEEEEERKGLLPALGALSGAQVSWPQLLSTYRLLSLSELVLSRAVRELL 222  
  
QY 205 LKK 207  
DB 223 LSK 225

RESULT 14  
US-09-810-052-2  
; Sequence 2, Application US/09810052  
; Patent No. US20020009775A1  
; GENERAL INFORMATION:  
; APPLICANT: Conklin, Darrell C.  
; APPLICANT: Presnell, Scott R.  
; TITLE OF INVENTION: HELICAL PROTEIN ZALPHA51  
; FILE REFERENCE: 00-24  
; CURRENT APPLICATION NUMBER: US/09/810,052  
; CURRENT FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/190,410  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/199,443  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 232  
; TYPE: PR1  
; ORGANISM: Homo sapiens  
US-09-810-052-2

Query Match 7.0%; Score 86; DB 10; Length 232;  
Best Local Similarity 22.7%; Pred. No. 0.42;  
Matches 50; Conservative 25; Mismatches 95; Indels 50; Gaps 7;  
  
QY 19 LWHLPAVPALNRTGDPGPGPSIQK-----TYDLTRYLEHQLRSLAGTYLN----- 63  
DB 15 VWGFRPPG-----RPQLSLQELRREFTVSLHLARLSEVRGQAHRAESHLPGVNL 67  
  
QY 64 YLGPPFNEPDNPPRLGAETLPRATVNLVWRSNDLRRLRLTQNYEAYSHLLCYLRGLNRQ 123  
DB 68 YLLP-----LG-EQLPDVSLTFQAWRLSDPRLCFISTTLQPPHAPLGLGTQ 115  
  
QY 124 AATAELRR-SLAHFTCSLQGLGSLAGTYMATLGYPLPOP-----LPGTEP 167

Sun Feb 2 08:32:33 2003

Db 116 GRWTNMRMQLWAMRLDLRLQRLRQVLAAGFNLPDEEEEBEERKGLPGALG 175  
Qy 168 AWAGCPAHSDFLOKMDDFWLLKELQTLWRSKADFNRLKK 207  
Db 176 SALQGPAQVSWPQLSTYRLHLSLELVLSRAVRELLLSK 215

RESULT 15  
US-09-730-617-2  
; Sequence 2, Application US/09730617  
; Patent No. US20020068279A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Prayaga, Sudhirdas K  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Mezes, Peter S  
; TITLE OF INVENTION: No. US20020068279A1el Proteins and Nucleic Acids Encoding the Sam  
; FILE REFERENCE: 15966-609  
; CURRENT APPLICATION NUMBER: US/09/730,617  
; CURRENT FILING DATE: 2000-12-05  
; PRIOR APPLICATION NUMBER: 60/169,056  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 60/169,886  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/169,866  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/170,252  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/175,740  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-730-617-2

Query Match 6.7%; Score 81.5; DB 10; Length 348;  
Best Local Similarity 28.2%; Pred. No. 1.9;  
Matches 74; Conservative 19; Mismatches 88; Indels 81; Gaps 17;  
Qy 9 WGMACLCCTVLWHLPAVPALNRTGPGGPS-----IQKTYDLTRYLHQHLSLAGT 60  
Db 3 WSCLA--CTLLRVLPHVLSLUR--DPVPVPTGTLFHSCTSTNFCASFLEVAVEA-AGI 57  
Qy 61 YLNYLIG---PP-----FNEPDP---FNPRLGAETLPRATVNLVWRS--LN 98  
Db 58 TPWTVGSEHPCCPYPSLHASPTDSFNRPSPAPLNRPSAGE--PRT---EAFPSGLK 111  
Qy 99 DRLRLTQNYEAYSHLLCYLRLGNRQAATAELRRSLAHFCTSLQGLLGSAGWATLGYPL 158  
Db 112 ARVGGTILAEA-----GLNSQGHAVEPVPS-----GPSGSKGCVLIKGRPS 153  
Qy 159 PQPLP-----GTEPAWAPGPAHSDFLOKMDDFWLLK-----ELQTLWRSKAD 201  
Db 154 RMPKARECPVDRENLLLTNPVPSLLQLLSSSPFCIKVETEQRSNAEFDLQS---RAARD 210  
Qy 202 FN-RLKKKM-QPPAASVTIHL 221  
Db 211 YNSRLLLKLGQIPAAKGSFLE 232

Search completed: January 27, 2003, 15:49:55  
Job time : 10 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:49:37 ; Search time 23.5426 Seconds  
(without alignments)  
6508.980 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 1458

Sequence: 1 attaaagcttcgcggagcc.....tctctcttcgtccccccc 797

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cn2 1/USPTO spool/US09931704/runat 27012003 154126 3648/app query.fasta\_1.7189  
-DB=PIR 73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09931704 @CN 1 1 351 @runat 27012003 154126 3648 -NCPU=6 -ICPU=3  
-NO XLPXY -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR 73:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	165	11.3	1532	2	A61262 collagen alpha 1(X
2	161	11.0	623	2	A45137 collagen alpha 4(I
3	160	11.0	383	2	S32975 Gene BCRF2 protein
4	159	11.0	660	1	QBE3 BHLF1 protein - hu
5	158	10.8	574	2	T32556 Wiskott-Aldrich sy
6	158	10.8	574	2	T38819 wiskott-aldrich sy
7	157	10.8	1843	2	S18803 collagen alpha 1(V
8	156.5	10.7	325	2	T32248 hypothetical prote
9	156	10.7	1690	1	CGHU1B collagen alpha 4(I
10	155.5	10.7	319	2	T32250 hypothetical prote
11	154.5	10.6	453	2	S18804 collagen alpha 4(I
12	152	10.4	382	2	S20375 collagen alpha 3(V
13	152	10.5	627	2	A44112 spidroin 2, dragli
14	151.5	10.4	1838	1	CGHU1V collagen alpha 1(V

15	150.5	10.3	302	2	T15936 hypothetical prote
16	150.5	10.3	418	2	T15142 hypothetical prote
17	150	10.3	278	2	B39066 proline-rich prote
18	149.5	10.3	1433	2	A46053 bullous pemphigoid
19	149	10.2	1146	2	A38587 collagen, cornea-s
20	148.5	10.2	1670	1	CGHU3B collagen alpha 3(I
21	147.5	10.1	317	2	T19143 hypothetical prote
22	147.5	10.1	684	2	A53019 collagen alpha 1(X
23	146	10.0	688	2	A53330 collagen alpha 2(I
24	145.5	10.0	304	2	T22602 hypothetical prote
25	145.5	10.0	317	2	A28996 proline-rich prote
26	144.5	9.9	240	2	A24264 proline-rich prote
27	144.5	10.0	385	2	T20410 hypothetical prote
28	144.5	9.9	1137	2	A86335 T20H2.9 protein -
29	144	9.9	1466	1	CGHU7L collagen alpha 1(I
30	143.5	9.8	317	2	T29960 hypothetical prote
31	143.5	9.8	435	2	T15143 hypothetical prote
32	143.5	9.9	481	2	A35628 loricerin - mouse
33	143	9.8	255	2	JQ0320 PTB-associated spl
34	143	9.8	1006	2	T42731 atrophin-1 related
35	143	9.8	1049	1	CGBO7S collagen alpha 1(I
36	143	9.8	3149	1	QBE8 BPLF1 protein - hu
37	142.5	9.8	240	2	B24264 proline-rich prote
38	142.5	9.8	505	2	S72273 actin-depolymerizi
39	142	9.7	286	2	S34665 collagen, cuticula
40	142	9.7	707	2	A46302 PTB-associated spl
41	141.5	9.7	227	2	C29149 proline-rich prote
42	141.5	9.7	403	2	S52796 prpl2 protein - hu
43	141.5	9.8	1215	2	T32734 myosin-1A - Acanth
44	141	9.7	1201	2	G86441 unknown protein [i
45	140	9.7	839	2	F75518 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

A61262  
collagen alpha 1(XVII) chain - human (fragment)  
N:Alternate names: bullous pemphigoid 180K autoantigen BPAG2; bullous pemphigoid antigen  
C:Species: Homo sapiens (man)  
C:Date: 12-May-1994 #sequence revision 12-Jul-1996 #text\_change 21-Jul-2000  
C:Accession: I56325; I55345; A61262  
R:Giudice, G.J.; Emery, D.J.; Diaz, L.A.  
J. Invest. Dermatol. 99, 243-250, 1992  
A:Title: Cloning and primary structural analysis of the Bullous pemphigoid autoantigen  
A:Reference number: I56325; MUID:92381323; PMID:1324962  
A:Accession: I56325  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1532 <RES>  
A:Cross-references: GB:M91669; NID:g179516; PIDN:AAA35605.1; PID:g179517  
R:Li, K.H.; Sawamura, D.; Giudice, G.J.; Diaz, L.A.; Mattei, M.G.; Chu, M.L.; Uitto, J  
J. Biol. Chem. 266, 24064-24069, 1991  
A:Title: Genomic organization of collagenous domains and chromosomal assignment of human  
A:Reference number: I55345; MUID:92084712; PMID:1748679  
A:Accession: I55345  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 543-890, 'P' <RE2>  
A:Cross-references: GB:M63730; NID:g179520; PIDN:AAA51839.1; PID:g179521  
R:Giudice, G.J.; Squiquera, H.L.; Elias, P.M.; Diaz, L.A.  
J. Clin. Invest. 87, 734-738, 1991  
A:Title: Identification of two collagen domains within the bullous pemphigoid autoantigen  
A:Reference number: A61262; MUID:91123476; PMID:1846881  
A:Accession: A61262  
A:Molecule type: mRNA  
A:Residues: 543-890, 'P' <GIU>  
C:Genetics:  
A:Gene: GDB:COL17A1; BPAG2; BP180  
A:Cross-references: GDB:I31396; OMIM:113811  
A:Map position: 10q24.3-10q24.3

Alignment Scores:

Pred. No.:	0.000125	Length:	1532
Score:	165.00	Matches:	77
Percent Similarity:	29.41%	Conservative:	23
Best Local Similarity:	22.85%	Mismatches:	96
Query Match:	11.32%	Indels:	144
DB:	2	Gaps:	13

US-09-931-704-1 (1-797) x A61262 (1-1532)

Qy	20	CGGGCTCGCCCTCCCACTCCGCCAGCTCCGGAGAGAGAGCGCACCCGGCGCGCCAC	G 79
Db	667	ArgGlyGluAlaGlyProProGlySerGlyGluLysGlyGluArgGlyAlaAlaGlyGlu	686
Qy	80	CCCCAGCCCATGACCTCCGAGC	103
Db	687	ProGlyProHisGlyProProGlyValProGlySerValGlyProLysGlySerSerGly	706
Qy	104	-----AGGGGACTCGTGGGGATGTT	124
Db	707	SerProGlyProGlnGlyProProGlyProValGlyLeuGlnGlyLeuArgGlyGluVal	726
Qy	125	AGCGTGCT-----GTGCACGGTCTCGACCTCCCTGCGAGTCCGAGCTCT	172
Db	727	GlyLeuProGlyValLysGlyAspLysGlyProMetGlyProProGlyProLysGlyAsp	746
Qy	173	CAATCGCACAGGACCCAGGGCTGGCCCCCTCATCCGAGAAACCTATGACCTCACCG	232
Db	747	GlnGlyGluLysGlyProArgGlyLeuThrGlyGluProGly-----MetArgGly	763
Qy	233	CTACTCGAGACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTCGGGCCCCC	292
Db	764	LeuProGlyAla-----ValGlyGluPro	771
Qy	293	TTTCAACGAGCCAGACTTCAACCTCCCGCGCTGGGGGAGAGACTCTGCCACAGGCCAC	352
Db	772	GlyAlaLysGlyAlaMetGlyProAlaGlyProAspGlyHisGlnGlyProArgGlyGlu	791
Qy	353	TGTTGACTTGGAGTGTGGCGAGCCTCAATGACAAACTCGCGCTGACCCAGAACTACA	412
Db	792	GlnGlyLeuThrGlyMet-----ProGlyLeuArg	801
Qy	413	GGCTACAGCCACCTTCTGTGTTACTTTCGTGCGCTCAACCGTCAGGCTGCCACTGTA	472
Db	802	Gly-----ProProGly-----ProSerGly-----	808
Qy	473	GCTGGCGGAGCCTGGC---CACTTCTGCAACAGCTCCAGGCTCGTGGG---	523
Db	809	-----AspProGlyLysProGlyLeuThrGlyProGlnGlyProGlnGlyLeuPro	825
Qy	524	CAGCATTCGGGGCTCATGGAGCTCTGGGTACCCACTGCCCGCCAGCGCTGCTGGGAC	583
Db	826	GlyThrProGlyArgProGlyLysGlyGluProGlyAlaPro-----	840
Qy	584	TGAACCCACTTGGACTCTCGGCCCTGCCACAGTACTTCTCTCCAGAAGATGACGACTT	643
Db	841	-----GlyLysIle	843
Qy	644	CTGGCTGCTGAAGAGCTGACAGCTGGCTGTG-----GCG	679
Db	844	ValThrSerGluGlySerSerMetLeuThrValProGlyProProGlyProGlyAla	863
Qy	680	CTCGCCCAAGGACTTCAACCGCTCAAGAGAGATGACGACCTCCAGCAGCTGC	733
Db	864	MetGlyProProGlyProGlyAlaProGlyProAlaGlyProAlaGlyLeuProGly	883
Qy	733	-----	733
Db	884	HisGlnGluValLeuAsnLeuGlnGlyProProGlyProProGlyProArgGlyProPro	903
Qy	734	-----AGT	736
Db	904	GlyProSerIleProGlyProProGlyProArgGlyProProGlyGluGlyLeuProGly	923

```

Qy 737 CACCTGCACCTGGGGGCTCATGGCTTCGACTCTTCGACCTTCTCCTCTGCTGCTCCCCC 793
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 924 ProProGlyProProGlySerPheLeuSerAsnSerGluThrPheLeuPheGlyProPro 943
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
A45137
collagen alpha 4(IV) chain - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-1999-1994
C/Accession: A45137
R/Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A/Title: Isolation and sequencing of cDNAs encoding the alpha 1(I) chain of
A/Reference number: S28777; MUID:93054733; PMID:14229714

```

Alignment Scores:	
Pred. No.:	0.000246
Score:	161.00
Percent Similarity:	36.51%
Best Local Similarity:	30.71%
Query Match:	11.04%
DB:	2
Length:	623
Matches:	74
Conservative:	14
Mismatches:	75
Indels:	78
Gaps:	16

US-09-931-704-1 (1-797) x A45137 (1-623)

Qy	19	CCGGGGCTCGCGCTCTCCCATCTCCGCCAGCCTCTCGGGAGAGAG-	-----CCGCACCCGGCC	72
Db	183	ProIysGlyAepIleProAspProGlyProProGlyAspGlnGlyLeuProGlyProAsp	:     :     :     :     :     :	202
Qy	73	GGCCACGCCCCAGCCCATCGACTCCGAGC	:     :     :     :     :     :	108
Db	203	GlyProArgGlyAlaPro-GlyProAlaGlyProProGlySerValAspLeuLeuYsGI	:     :     :     :     :     :	222
Qy	109	ACTCGTGGGGGATGTAGCGTCTGTGCACGGTGCTCTGGCACCTCCC	:     :     :     :     :     :	157
Db	222	YgluProGlyAspCysGlyValPro-----GlyProProGlyProProGlyProProGI	:     :     :     :     :     :	240
Qy	158	-----TGACGTGCC-----ACGCTCTCAATCGCACAGGGGCCAC	:     :     :     :     :     :	192
Db	240	yProProGlyCysGlnGlyValProGlyCysAspGlyHisAspGlyGlnLysGlyProMe	:     :     :     :     :     :	260
Qy	193	GGCGCTGGCCC--CTCCATCCAGAAAACCTATGACCTCACCGCTACCTCGAGACCAAC	:     :     :     :     :     :	249
Db	260	tGlyPheProGlyLeuGlnGly-----ProHisGlyLeuLeuProGlyLeuProGI	:     :     :     :     :     :	276
Qy	250	TCCGACGCTTGGCTGGGACCTATCTGAACCTACTGGGGCCCCCTTTCAACGAGCC	:     :     :     :     :     :	304
Db	276	yGluLys-----GlyLeuProGlySerProGlyArgLysGlyProth	:     :     :     :     :     :	290
Qy	305	-----AGACTTCAACCTCCCGCTCGGGGCGAGACACTCTCCCGCAGG	:     :     :     :     :     :	348
Db	290	rGlyProProGlyTyrrArgGlyGluProGlyProProAlaAspGluAspSerCysProAr	:     :     :     :     :     :	310
Qy	349	CCACTGTTGACTTGGAGGTGTGGGAAGCCTCAATGACAAACTGCGGCTGACCCAGAACT	:     :     :     :     :     :	408
Db	310	gIleProGlyLeuProGlyValProGlyPro-----	:     :     :     :     :     :	320
Qy	409	ACGAGGGCTACAGCCACTTCTGTGTTACTTGTGGTGGCTCT	:     :     :     :     :     :	459
Db	321	-ArgGly-----ProGluGlyAlaMetGlyPheProGlyGlnArgGlyProProGI	:     :     :     :     :     :	337
Qy	460	CTGCCACTGTGACTCGCGC-----	:     :     :     :     :     :	489
Db	337	yGlnGlyCysLysGlyGluProGlyLeuAspGlyLysArGlyArgAspGlyValProGI	:     :     :     :     :     :	357
Qy	490	CCCACCTCTGCACAGCGCTCCAGGGCTGTGTGGGACGACTTGGCGGCGTCAATGGCAGCTC	:     :     :     :     :     :	549



QY 540 TGACGCCCGCAATCTGCTCCACAGAGCCCTGGAGGCTGGTGCAGAGTGGG----- 490  
 Db 246 -----GlnArgCysProAlaGlyProProPro---ThrArgSerGlyAlaAlaA 261  
 QY 489 -----CAGGCTGC-----GGCGCAGCT 472  
 Db 262 GlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysPro 281  
 QY 471 CAGCAGTGGCAG-----CTGACGGTGTAGGCCAC 442  
 Db 282 ArgThrTrpArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaGlyGln 301  
 QY 441 GCAAGTAACACAGAGTGCTGTAGCCCTCGTAGT----- 406  
 Db 302 ArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrProAla 321  
 QY 405 -----TCTGGTCAGCGCGAGTTGTCAITGAGGCTTCGCCACACCTCCAGT 358  
 Db 322 AlaProGlyProGlyGlyAlaAlaVal-ProSerGlyAlaThrProHisProGluAr 341  
 QY 357 CAACAGTGGCCCTGGCGAGAGTCTCTGCCCCACAGCGGGGAGGCTGAAGTCTGGCTCGT 298  
 Db 341 gGlySerGlyProAlaAspPro-----ProAlaAlaAlaArgLeuProProGluArgL 359  
 QY 297 TGAA-----AGGGGGGCCAGGTAGTTCAGATAGTCCACGCCA 259  
 Db 359 nGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysProAlaGlyPro--Pro 378  
 QY 258 AGCTGGGAGTTGGTCTCAGGTAGCGGTGAGGTCAATAGTTTCTGGATGAGGGGC 199  
 Db 379 ProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArg 398  
 QY 198 CAGGCCCTGGGTCCCTGTGCGATTGAGAGCTGGCACTGGCAGGAGGTGCCAGAGCACCG 139  
 Db 399 SerAlaArgAsnProGlyCysPro-ArgThrTrpArgArgSer----- 413  
 QY 138 TGCACAGGCACGTAAATATCCCCACAGTCCCTGTCTCGAGGTCCATGGGGCTGGGGC 79  
 Db 414 -GlyAlaGlnArgGlyHisProPro-ProGlyAlaGlyGlnArgProSerGlyProThrG 433  
 QY 78 TGGGCGGG-----CCGGTGGCGCTCTCTCCCGAGGCTGGCGAGTGGGAGTGGAGGGC 28  
 Db 433 lyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyG 453  
 QY 27 GAGCGCGCGCTCCGGCG 11  
 Db 453 lyAlaAlaValProSer 458

RESULT 5  
 T43556  
 Wiskott-Aldrich syndrome protein homolog - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
 C:Accession: T43556  
 R:Zankel, T.C.; Ow, D.W.  
 A:Submitted to the EMBL Data Library, December 1997  
 A:Description: A Wiskott-Aldrich Syndrome protein homolog in Schizosaccharomyces pombe,  
 A:Reference number: 222575  
 A:Accession: T43556  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-574 <ZAN>  
 A:Cross-references: EMBL:AF038575; PIDN:AAB92587.1  
 A:Experimental source: strain JS21  
 C:Genetics:  
 A:Gene: wapl  
 A:Map position: I  
 A:Introns: 72/3; 519/3; 564/1  
 Alignment Scores:  
 Pred. No.: 0.000412 Length: 574  
 Score: 158.00 Matches: 77  
 Percent Similarity: 34.74% Conservative: 22

US-09-931-704-1 (1-797) x T43556 (1-574)  
 QY 11 GCCTGAGCCCGGCTCCGCC-----TCCACTCCGCGCAGCTCCCGGAG 55  
 Db 319 ArgArgAsnArgGlyLysProProIleGlyAsnGlySerSerAsnSerSerLeu----- 336  
 QY 56 AGGAGCCGACCCCGCGCCCGCCAGCCCGCCCTGGACCTCCGAGCAGGGGACTCGTG 115  
 Db 337 -----ProProProProProProProProProProProProProProProPro 348  
 QY 116 GGGGATGTAGCGTCTGTGCACGCTGCTCGGCA-----CCTCCCTGCAGT 163  
 Db 349 GlySerIleProLeuProProGlnGlyArgSerAlaProProProProProProProPro 368  
 QY 164 GCCAGCTCTCAATCGCACAGGGGACCCAGGCGCTGGCCCTCCATCCAGAAAACCTATGA 223  
 Db 369 AlaProSerThrGlyArgGlnProPro-----ProLeuSerSerSerArgAlaVal 385  
 QY 224 CTTACCCG-----CTACCTGGAGCACCACTCCGCGAGCTTGCTGG 265  
 Db 386 SerAsnProProAlaProProProAlaIleProGlyArgSerAlaProAlaLeu----- 403  
 QY 266 GACCTATCTGAATACCTTGGGCGCCCTTTCAACGAGCCAGACTTCAACCTCCCGCCT 325  
 Db 404 -----ProProLeuGlyAsnAlaSerArgThrSerThrProPro 416  
 QY 326 GGGGACAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGA 385  
 Db 417 -----ValProThrProProSerLeu 423  
 QY 386 CAATCTGGCTGACCCAGAACTACAGGCCTACAGCCACTCTCTGTACTTGGTGG 445  
 Db 424 ProProSerAlaProProSerLeuProProSerAlaProProSerLeuProMetGlyAla 443  
 QY 446 CTTCAACCGTCAGGCTGCCACTGTCTGAGCTGGCGCGCAGCTGGCCCTTCTGCACCAG 505  
 Db 444 ProAla-----AlaProProLeuProProSerAlaProIleAla 456  
 QY 506 CTTCCAGGCTCTGTGGCAGCATTTGGGGGCTCATGGCAGCTCTGGCTTACCATCTGCC 565  
 Db 457 ProProLeuProAla-----GlyMetProAlaAla 466  
 QY 566 CAGCGCTGCTGGGACTGAACCCACTTGGACTCTGGCTGCCCTGCCACAGTCA----- 619  
 Db 467 ProProLeuProProAlaProAlaProProProProProProProProProProAla 486  
 QY 620 -----CTTCTCCAGAAATGACGACTT----- 643  
 Db 486 aProValAlaSerIleAlaGluLeuProProGlnGlnAspGlyArgAlaAsnLeuMetAlaSe 506  
 QY 644 -CTGGCTCTGAAGAGCTGCAGACTGGCTGTGGCGCTTCGGCCAAAGGACTTCAACCGGC 702  
 Db 506 rIleArgAlaSerGlyGlyMetAspLeuLeuLysSerArgLysValSerAlaSerProSe 526  
 QY 703 TCAGAAGAAGATGACGCTCCAGCAGCTGCAGTCACTCCCT-----GCACCTGGGG 753  
 Db 526 r-----ValAlaSerThrLysThrSerAsnProProValGluAlaProProSe 542  
 QY 754 CTCATGGCTTCG 766  
 Db 542 rAsnAsnLeuMet 546

## RESULT 6

T38819

wiskott-aldrich syndrome protein homolog 1 - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T38819  
 R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.



QY 454 GTGAGGTGCCACTGTGAGTGGCGG-----CAGCTGGCCCACTT-----CTGCACCA 504  
Db 1186 oThrGlyProGln-----GlyProLleGlyGlnProGlyProSerGlyAlaAspG1 1203  
QY 505 GCCTCCAGGCGCTGCTGGCGAG-----CATTGCGGCGCTCATGGCAGCTC 549  
Db 1203 yGluProGlyProArgGlyGlnGlnGlyLeuPheGlyGlnLysGlyAspGluGlySerAr 1223  
QY 550 TGGGCTACCACTGCCCGCCGCGCTGCTGGGAGTGAACCACTTGGACTCTGGCCCTG 609  
Db 1223 gGlyPheProGlyProProGlyPro-----ValGlyLeuGlnGlyLeuProG1 1239  
QY 610 CCCACAGTACTTCTCCAGAGATGGA 637  
Db 1239 yPro-----ProGlyGluLysGly 1245  
RESULT 8  
T32248  
hypothetical protein T15B7.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
C:Accession: T32248  
R:Pauley, A.; Gattung, S.  
Submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid T15B7.  
A:Reference number: Z31139  
A:Accession: T32248  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-325 <PAU>  
A:Cross-references: EMBL:AF022985; PIDN:AAB69961.1; GSPDB:GN00023; CESP:T15B7.4  
A:Experimental source: strain Bristol N2; clone T15B7  
C:Genetics:  
A:Gene: CESP:T15B7.4  
A:Map position: 5  
A:Introns: 266/1  
C:Superfamily: unassigned collagens  
Alignment Scores:  
Pred. No.: 0.000529 Length: 325  
Score: 156.50 Matches: 66  
Percent Similarity: 35.19% Conservative: 10  
Best Local Similarity: 30.56% Mismatches: 86  
Query Match: 10.73% Indels: 54  
DB: 2 Gaps: 10  
US-09-931-704-1 (1-797) x T32248 (1-325)  
QY 56 AGGAGCGCAGCCCGCGCGCCAGCC-----CAGCCCATGGACCTCCG 100  
Db 85 GlnSerArgGlyCysProAlaGlyProProGlyProProGlyGlnProGlyAlaProGly 104  
QY 101 AGCAGGGGACTCGTGGGGGATGTTAGCTGCTGTGCGACGGTGTCTGGCCTCC--- 157  
Db 105 GluGlnGlyHisProGlyLeuAlaGlyGlnProGlySerGlyAlaArgIleAsnProAla 124  
QY 158 -----TGCAGTGCAGCTCTCAATCGCACGGGACCC 190  
Db 125 ThrGlyArgProGlyPheCysIleThrCysProAlaGlyAlaProGlyProAlaGlyPro 144  
QY 191 AGGGCTGGCCCTCCATCCAGAAACCTATGACCTCCTACCTCGCTACCTGGACACCACT 250  
Db 145 -----ProGlyAlaProGly 149  
QY 251 CCGCAGCTTGGTGGAGCTAFTCTGAACCTACTGGGCCCCCTTTCAAGA---GCCAGA 307  
Db 150 ProLys-----GlyAsnAsnGlyGlnProGlyAlaProAlaGlnSerGlyGlyArg 166  
QY 308 CTTCAACCTCCCGCCTGG-----GGCAGAGACTCTGCCAGGGCCAC 352  
Db 167 GlyProProGlyProArgGlyProAlaGlyAspAlaGlySerProGlyGlnProGlyHis 186  
QY 353 TGTTCAGTCTGGAGGTGTGGCGAAGCCTCAATGATCAAACTGCGGCTGACCCAGAACTACGA 412

Db 187 -----ProGlySerProGlyAsnProGlyArgGlyGlyGlnArgSerArg 201  
QY 413 GGCCTACAGCCACCTTCTGTATTCTGGTGGCTCAACCGTCAGGC----- 460  
Db 202 GlyLeuProGlyProSerGlyArgProGlyProProGlyProAlaGlyGlyProGlyGln 221  
QY 461 ---TCCCACTGTGAGCTCGCGCAGCTGGCCCACTTCTGCACCAAGCTCCAGGCGCT 517  
Db 222 ProGlyHisSerGlyGlyAlaGlySerProGlyPro-----GlnGlyProProGlyPro 239  
QY 518 GCTGGGCGAG-----CATTGCGGCGCTCATGGCAGCTCTGGGTACCCACTGCCCGCC 571  
Db 240 SerGlyGlnProGlyHisSerGlyAsnAspGlyValProGlyAlaProGlyAsnProGly 259  
QY 572 GCTGCTGGGAGTGAACCCACTTGGACTCTGGCCCTGCCACACAGT 617  
Db 260 Ser-ProGlyGlyAspAlaAlaTyrCysProCysProAlaArgSer 274  
RESULT 9  
CGHULB  
collagen alpha 4(IV) chain precursor - human  
N:Alternate names: procollagen alpha 4(IV) chain  
C:Species: Homo sapiens (man)  
C>Date: 06-Feb-1995 #sequence\_revision 03-Oct-1995 #text\_change 16-Jun-2000  
C:Accession: A55360; S36854; S28777  
R:Leinonen, A.; Marlyama, M.; Mochizuki, T.; Tryggvason, K.; Reiders, S.T.  
J. Biol. Chem. 269, 26172-26177, 1994  
A:Title: Complete primary structure of the human type IV collagen alpha4(IV) chain.  
A:Reference number: A55360; MUID:95014445; PMID:7523402  
A:Accession: A55360  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1690 <LEI>  
A:Cross-references: GB:X81053; NID:G574805; PIDN:CAA56943.1; PID:9574806  
R:Sugimoto, M.; Ohashi, T.; Yoshioka, H.; Matsuo, N.; Ninomiya, Y.  
FEBS Lett. 330, 122-128, 1993  
A:Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen  
A:Reference number: S36854; MUID:93374047; PMID:8365481  
A:Accession: S36854  
A:Molecule type: DNA; mRNA  
A:Residues: 1219-1658, 'FE', 1661-1690 <SUG>  
A:Cross-references: DBJ:DJ1391; NID:9440365; PIDN:BAA04214.1; PID:9457161  
A:Experimental source: whole eye  
R:Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.  
J. Biol. Chem. 267, 23753-23758, 1992  
A:Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain  
A:Reference number: S28777; MUID:93054733; PMID:1429714  
A:Accession: S28777  
A:Molecule type: DNA  
A:Residues: 1407-1424, 'G', 1426-1430, 'A', 1432-1439, 'L', 1441-1507 <KAM>  
A:Cross-references: GB:L01475; GB:L01476  
A:Note: the codons given for 1438-Asp (GAG) and 1443-Gly (GCA) are inconsistent with  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C:Genetics:  
A:Gene: GDB:COL4A4  
A:Cross-references: GDB:132673; OMIM:120131  
A:Map position: 2q35-2q37  
A:Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete  
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands w  
C:Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha  
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimer  
er associations in the interrupted helical domain (with disulfide and desmosine cross  
C:Function:  
A:Description: minor structural component of extracellular basement membrane in kidney  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydro  
F;1-38/Domain: signal sequence #status predicted <SIG>  
F;39-1690/Product: collagen alpha 4(IV) chain #status predicted <MAT>  
F;39-61/Domain: amino-terminal nonhelical, NHI <NHI>  
F;62-1466/Region: interrupted helical  
F;94-96/Region: cell attachment (R-G-D) motif

















CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L01477; -; NOT\_ANNOTATED\_CDS.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001442; ProcollagenC4.  
 DR Pfam; PF01391; Collagen; 5.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagenC4; 2.  
 DR SMART; SM00111; C4; 2.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
 FT NON\_TER 1 1  
 FT DOMAIN <1 392 TRIPLE-HELICAL REGION.  
 FT DOMAIN 393 623 NONHELICAL REGION (NC1).  
 FT DISULFID 413 502 OR 499 (BY SIMILARITY).  
 FT DISULFID 446 499 OR 502 (BY SIMILARITY).  
 FT DISULFID 458 464 BY SIMILARITY.  
 FT DISULFID 521 619 OR 616 (BY SIMILARITY).  
 FT DISULFID 555 616 OR 619 (BY SIMILARITY).  
 FT DISULFID 567 574 BY SIMILARITY.  
 SQ SEQUENCE 623 AA; 62393 MW; C4C9B31242FE82 CRC64;

Alignment Scores:  
 Pred. No.: 0.00468 Length: 623  
 Score: 161.00 Matches: 74  
 Percent Similarity: 36.51% Conservative: 14  
 Best Local Similarity: 30.71% Mismatches: 75  
 Query Match: 11.04% Indels: 78  
 DB: 1 Gaps: 16

US-09-931-704-1 (1-797) x CA44\_RABIT (1-623)

QY 19 CCGGGTCGCGCTCCACTCCGCGAGCTCCGGAGAGGAG-----CCGACCCCGGCC 72  
 DB 183 ProlysglyaspilProaspProGlyProGlyaspGinglyLeuProGlyProasp 202  
 QY 73 GGCCAGCCCGAGCCCATGACCTCCGAGC-----AGGGG 108  
 DB 203 GlyProargglyAlaPro-GlyProAlaGlyProGlySerValAspLeuLeuysG 222  
 QY 109 ACTGCTGGGGATTTAGCGTCTGTGACAGGTGCTTGGCAGCTCC----- 157  
 DB 222 yGluProGlyAspCysGlyValPro-----GlyProGlyProGlyProGlyProG 240  
 QY 158 -----TGCAGTGC-----AGCTCTCAATCGCAGGGGACCCAG 192  
 DB 240 yProProGlyCysGinglyValProGlyCysAspGlyHisAspGlyGinlysglyPro 260  
 QY 193 GGCCTGGCC---CTCCATCCAGAAACCTATGACCTCACCGCTACCTGGAGCACCAAC 249  
 DB 260 tGlyPheProGlyLeuGingly-----ProHisGlyLeuProGlyLeuProG 276  
 QY 250 TCCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAACAGGCC----- 304  
 DB 276 yGluLys-----GlyLeuProGlySerProGlyArgLysGlyProth 290  
 QY 305 -----AGACTTCAACCTCCCGCGCTGGGGCAGAGACTCTGCCAGGG 348  
 DB 290 rGlyProProGlyTyArgGlyGluProGlyProProAlaAspGluAspSerCysProAr 310  
 QY 349 CCACTGTTGACTTGGAGTGTGGGAGCGCTCAATGACAAACTGGCGGTGACCCAGAACT 408  
 DB 310 gIleProGlyLeuProGlyValProGlyPro----- 320  
 QY 409 ACGAGGCTACAGCCACTCTGTGTTACTTGGCGTGCCT-----CAACCGTCAGG 459

DB 321 -ArgGly-----ProGluGlyAlaMetGlyPheProGlyGlnArgGlyProProG 337  
 QY 460 CTGCCACTGCTGAGCTGGCCCG-----CAGCCTGG 489  
 DB 337 yGlnGlyCysLysGlyGluProGlyLeuAspGlyLysArgGlyArgAspGlyValProG 357  
 QY 490 CCCACTTCTGCACCAAGCTCCAGGCGCTGCTGGCAGCATTCGGCGCTCATGGCAGCTC 549  
 DB 357 y-----AlaProGlyProProGly---HisLysGlyAspThrGlyGluAl 371  
 QY 550 TGGGTACCCACTGCCCGCAGCGCTGCTGGGACTGAACCCACTTGGACTCTCCTGGGCCT 608  
 DB 371 aGlyArgProGlyAlaProGlyPro-ProGly-----ProThrGlyAspProGlyPro 388  
 RESULT 2  
 YHL1\_EBV STANDARD; PRT; 660 AA.  
 ID YHL1\_EBV  
 AC P03181;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Hypothetical BHLF1 protein.  
 OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 OX NCBI\_TaxID=10377;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84270667; PubMed=6087149;  
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,  
 RA Tufnell P.S., Barrell B.G.;  
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";  
 RL Nature 310:207-211(1984).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; V01555; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A03742; Q0B3.  
 KW Hypothetical protein; Early protein; Repeat.  
 FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.  
 FT REPEAT 149 273 1.  
 FT REPEAT 274 398 2.  
 FT REPEAT 399 523 3.  
 FT REPEAT 524 648 4.  
 SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

Alignment Scores:  
 Pred. No.: 0.00621 Length: 660  
 Score: 159.00 Matches: 93  
 Percent Similarity: 32.52% Conservative: 14  
 Best Local Similarity: 28.27% Mismatches: 125  
 Query Match: 11.00% Indels: 97  
 DB: 1 Gaps: 13

US-09-931-704-1 (1-797) x YHL1\_EBV (1-660)

QY 786 AAGAGAGAGAGTTCAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGT 727  
 DB 160 ArgArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaGly----- 175  
 QY 726 CTGGAGCGCTCATCTCTTCTTTGAGCGGTGAAGTCTTTGGCGAGCGCCAGCCAGCAGG 667  
 DB 176 ---GlnArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThr 194  
 QY 666 TCTGAGCT-----CCTTCAGCAGCAGAGTCTGCTCA 634

Db 195 ProAlaAlaProGlyProGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisPro 214  
 QY 633 TCTTCTGGAGGAAGTCACTGTGGG-----CAGGCGCCAG 601  
 Db 215 GluArgGlySerGlyProAlaAspProProAlaAlaAlaAlaArgLeuProGluArgGln 234  
 QY 600 GAGTCAAGTGGTTCAGTCCAGGAGCGCTGGGGCAGTGGTGTAGCCAGAGCTGCCA 541  
 Db 235 GluProArgLeuProGlnAspLeuAlaAlaAla----- 245  
 QY 540 TGACGCCGCCCAATGCTGCCAGCAGCCCTGGAGGCTGGTGCAGAGTGGG----- 490  
 Db 246 -----GlnArgCysProAlaGlyProProPro-----ThrArgSerGlyAlaAlaAla 261  
 QY 489 -----CCAGCTGC-----GGCGCAGCT 472  
 Db 262 GlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysPro 281  
 QY 471 CAGCAGTGGCAG-----CCTGACGGTTGAGGCCACCTCCAGT 442  
 Db 282 ArgThrTrpArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyGln 301  
 QY 441 GCAAGTAACACAGAGGTGCTAGCCTCGTAGT----- 406  
 Db 302 ArgProSerGlyProThrGlyGlyArgProAlaAlaAlaProGlyAlaProGlyThrProAla 321  
 QY 405 -----TCTGGGTCCAGCCGAGTGTTCATTGAGGCTTCGCCACACCTCCAGT 358  
 Db 322 AlaProGlyProGlyGlyAlaAlaVal-ProSerGlyAlaThrProHisProGluArg 341  
 QY 357 CAACAGTGGCTGGCGCAGAGTCTCTGCCCGCAGGCGGAGGTTGAAGTCTGGCTCGT 298  
 Db 341 gGlySerGlyProAlaAspPro-----ProAlaAlaAlaArgLeuProProGluArgGln 359  
 QY 297 TGAA-----AGGGGGCCCGAGGTAGTTCAGATAGTCCCGAGCCA 259  
 Db 359 nGluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyPro--Pro 378  
 QY 258 AGCTCGGAGTGTGCTCTCCAGTCCAGTCCGCTGAGGTTCATGATGTTCTCGATGGAGGGC 199  
 Db 379 ProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArg 398  
 QY 198 CAGGCGCTGGTCCCTGTGCGATGAGAGTGGCTGCTGAGGAGGCTCCAGAGCAGCCG 139  
 Db 399 SerAlaArgAsnProGlyCysPro-ArgThrTrpArgArgSer----- 413  
 QY 138 TGCACAGGACCGTAACATCCCGCAGAGTCCCTGCTCGAGGTCCAGGGCTGGGCTGGGC 79  
 Db 414 -GlyAlaGlnArgGlyHisProPro-ProGlyAlaGlyGlnArgProSerGlyProThrG 433  
 QY 78 TGGGCGCG-----CCGGGTGGCGCTCTCTCCGAGGCTGGCGAGTGGGAGGGC 28  
 Db 433 lGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyG 453  
 QY 27 GAGCGCGGCTCCGGC 11  
 Db 453 lAlaAlaValProSer 458  
 RESULT 3  
 CA44 HUMAN  
 ID CA44 HUMAN STANDARD; PRT; 1690 AA.  
 AC P53420;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 4(IV) chain precursor.  
 OS COL4A4.  
 GN Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;  
 RX MEDLINE=95014445; PubMed=7523402;  
 RA Lemmon A., Mariyama M., Mochizuki T., Tryggvason K., Readers S.T.;  
 RT "Complete primary structure of the human type IV collagen alpha 4(IV)  
 chain. Comparison with structure and expression of the other alpha  
 (IV) chains.";  
 RL J. Biol. Chem. 269:26172-26177(1994).  
 RN [2]  
 RP SEQUENCE OF 1-23 FROM N.A.  
 RX MEDLINE=98196854; PubMed=9537506;  
 RA Monota R., Sugimoto M., Ohashi T., Kigawa K., Yoshioka H.,  
 RNomiya Y.;  
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
 alpha4(IV) collagen chains are arranged head-to-head on chromosome  
 2q36.";  
 RL FEBS Lett. 424:11-16(1998).  
 RN [3]  
 RP SEQUENCE OF 1219-1690 FROM N.A.  
 RX TISSUE=Eye;  
 RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;  
 RT "cDNA isolation and partial gene structure of the human alpha 4(IV)  
 collagen chain.";  
 RL FEBS Lett. 330:122-128(1993).  
 RN [4]  
 RP SEQUENCE OF 1407-1507 FROM N.A.  
 RX MEDLINE=93054733; PubMed=1429714;  
 RA Kanagata Y., Mattei M.-G., Ninomiya Y.;  
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the  
 alpha 4 chain of basement membrane collagen type IV and assignment of  
 the gene to the distal long arm of human chromosome 2.";  
 RL J. Biol. Chem. 267:23753-23758(1992).  
 RN [5]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97338662; PubMed=9195222;  
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
 RT "The clinical spectrum of type IV collagen mutations.";  
 RL Hum. Mutat. 9:477-499(1997).  
 RN [6]  
 RP VARIANT AS SER-1201.  
 RX MEDLINE=95078927; PubMed=7987396;  
 RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,  
 RA Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,  
 RA Smeets H.J.M., Readers S.T.;  
 RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)  
 collagen genes in autosomal recessive Alport syndrome.";  
 RL Nat. Genet. 8:77-82(1994).  
 RN [7]  
 RP VARIANT BFH GUJ-897.  
 RX MEDLINE=96379660; PubMed=8787673;  
 RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,  
 RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;  
 RT "Benign familial hematuria due to mutation of the type IV collagen  
 alpha4 gene.";  
 RL J. Clin. Invest. 98:1114-1118(1996).  
 RN [8]  
 RP VARIANTS AS, AND VARIANTS.  
 RX MEDLINE=99011253; PubMed=9792860;  
 RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,  
 RA Cochard P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;  
 RT "Determination of the genomic structure of the COL4A4 gene and of  
 novel mutations causing autosomal recessive Alport syndrome.";  
 RL Am. J. Hum. Genet. 63:1329-1340(1998).  
 CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/  
 CC NIDOGEN.  
 CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE  
 CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,

COCHLEA, LUNG AND BRAIN.

-I- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.

-I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

-I- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.

-I- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF TYPE II AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME (AS), AN HEREDITARY GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE, HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN MALES AND FEMALES.

-I- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF FAMILIAL BENIGN HEMATURIA (FBH) OR THIN BASEMENT MEMBRANE DISEASE. FBH IS CHARACTERIZED BY PERSISTENT HEMATURIA, AN ELECTRON MICROSCOPICALLY DETECTABLE THIN GLOMERULAR BASEMENT MEMBRANE (GBM) AND AN AUTOSOMAL DOMINANT MODE OF INHERITANCE. RENAL FUNCTION REMAINS NORMAL. IN CHILDREN DIFFERENTIATION BETWEEN FBH AND AS CAN BE DIFFICULT, BECAUSE BOTH DISORDERS ARE MANIFESTED BY PERSISTENT HEMATURIA AND THIN GBM AT THAT AGE.

-I- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

-----

DR EMBL; X81053; CAA56943.1; -.  
DR EMBL; AB008496; BAA25065.1; -.  
DR EMBL; D17391; BAA04214.1; -.  
DR Genew; HGNC:2206; COL4A4.  
DR MIM; 120131; -.  
DR MIM; 141200; -.  
DR MIM; 203780; -.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR001442; Procollagnc4.  
DR Pfam; PF01391; Collagen; 21.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD000007; Collagen; 1.  
DR ProDom; PD003923; Procollagnc4; 2.  
DR SMART; SMO0111; C4; 2.  
KW Extracellular matrix; Connective tissue; Basement membrane; Repeat;  
KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;  
KW Polymorphism; Alport syndrome.  
FT SIGNAL 1 38  
FT CHAIN 39 1690 POTENTIAL.  
FT DOMAIN 39 64 COLLAGEN ALPHA 4(IV) CHAIN.  
FT FT 64 7S DOMAIN.  
FT FT 65 1459 TRIPLE-HELICAL REGION.  
FT FT 1460 1690 NONHELIICAL REGION (NC1).  
FT SITE 94 96 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 145 147 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 189 191 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 310 312 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 724 726 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 785 787 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 989 991 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 1206 1206 CLEAVAGE (BY COLLAGENASE)  
FT FT (BY SIMILARITY).  
FT SITE 1212 1214 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 1480 1569 OR 1566 (BY SIMILARITY).  
FT DSULFID 1513 1566 OR 1569 (BY SIMILARITY).  
FT DSULFID 1525 1531 BY SIMILARITY.  
FT DSULFID 1588 1686 OR 1683 (BY SIMILARITY).  
FT DSULFID 1622 1683 OR 1686 (BY SIMILARITY).  
FT DSULFID 1634 1641 BY SIMILARITY.



Db 1401 lyProGlyProGlyCysLysGlyGluProGlyLeuAspGlyArgArgGlyValAspG 1421  
 Qy 482 --CACCTGGCCCACTTTCGACCAAGCTCCAGGCGCTGCTGGGCGAGCATTCGGGGCTC 539  
 Db 1421 lyValProGly-----SerProGlyProGlyArgLys---GlyAspT 1435  
 Qy 540 ATGGCAGCTCTGGCTACCACTGCCAGCGCTGCTGGGAGTGAACCCACTTGGACT 599  
 Db 1435 hrGlyGluAspGlyTyrProGlyGlyProGlyPro-ProGly-----ProileGlyAsp 1452  
 Qy 600 CTGGCCCT 608  
 Db 1453 ProGlyPro 1455  
 RESULT 4  
 TRX2 HUMAN STANDARD; PRT; 2715 AA.  
 ID TRX2 HUMAN Q9UMN6; Q9UK25; Q9Y669; Q9Y668; Q15022; Q96GP2; Q96IP3;  
 AC 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Trithorax homolog 2 (Mixed lineage leukemia gene homolog 2 protein).  
 GN TRX2 OR HRX2 OR MLL2 OR MLL4 OR KIAA0304.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (LONG ISOFORM).  
 RA Angrand P.O., Valvane H., Jeanmougin F., Adamson A.,  
 van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,  
 Ramerlin J., Chambon P., Lessou R., Stewart A., Aasland R.;  
 RA "Mammalian trithorax- and ASH1-like proteins: putative chromatin  
 RT regulators which contain PHD fingers and SET domains";  
 RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE FROM N.A. (LONG ISOFORM).  
 RA Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K.,  
 Garcia E., Kyle A., Ramirez M., Stilwagen S., Garnes J., Danganan L.,  
 Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,  
 Carrano A.V.;  
 RA "Sequence analysis of a 1 Mb region in human 19q13.1";  
 RT Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RP SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).  
 RC TISSUE=Testis, and Leukocyte;  
 RX MEDLINE=20105772; PubMed=10637508;  
 RA Huntsman D.G., Chin S.-F., Mulleris M., Batley S.J., Collins V.P.,  
 Wiedemann L.M., Aparicio S., Caldas C.;  
 RA "MLL2, the second human homolog of the Drosophila trithorax gene, maps  
 RT to 19q13.1 and is amplified in solid tumor cell lines";  
 RL Oncogene 18:7975-7984(1999).  
 RN [4]  
 RP SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 Miyaajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RA "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RL code for large proteins in vitro";  
 RN DNA Res. 4:141-150(1997).  
 RN [5]  
 RP SEQUENCE OF 1918-2715 FROM N.A.  
 RC TISSUE=Brain, and Skin;  
 RA Strausberg R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).  
 RC TISSUE=Placenta, and Bone marrow;  
 RX MEDLINE=99339983; PubMed=10409430;  
 RA Fitzgerald K.T., Diaz M.O.;  
 RA "MLL2: A new mammalian member of the trx/MLL family of genes.";

Genomics 59:187-192(1999).  
 CC -!- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC TRUNCATED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.  
 CC ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,  
 CC SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL  
 CC BLOOD LYMPHOCYTES, AND PLACENTA.  
 CC -!- DISEASE: OFTEN AMPLIFIED IN PANCREATIC CARCINOMAS.  
 CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.  
 CC -!- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ007041; CAB45385.1; -;  
 DR EMBL; AD000671; -; NOT ANNOTATED\_CDS.  
 DR EMBL; AB002302; BAA20763.2; -;  
 DR EMBL; AF186605; AAD56420.1; -;  
 DR EMBL; AF104918; AAD17932.1; -;  
 DR EMBL; AF105279; AAD26113.1; -;  
 DR EMBL; BC009337; AAH09337.1; -;  
 DR EMBL; BC007353; AAH07353.1; -;  
 DR EMBL; AF105280; AAD26112.1; -;  
 DR MIM; 606834; -;  
 DR InterPro; IPR000637; AT hook.  
 DR InterPro; IPR003889; FYrich\_C.  
 DR InterPro; IPR003888; FYrich\_N.  
 DR InterPro; IPR003616; PostSET.  
 DR InterPro; IPR001214; SET.  
 DR InterPro; IPR002857; Znf\_CXXC.  
 DR InterPro; IPR001965; Znf\_PHD.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00628; PHD; 3.  
 DR Pfam; PF00856; SET; 1.  
 DR SMART; SM00384; AT hook; 1.  
 DR SMART; SM00542; FYRC; 1.  
 DR SMART; SM00541; FYRN; 1.  
 DR SMART; SM00249; PHD; 4.  
 DR SMART; SM00508; PostSET; 1.  
 DR SMART; SM00184; RING; 1.  
 DR SMART; SM00317; SET; 1.  
 DR PROSITE; PS0280; SET; 1.  
 DR PROSITE; PS01359; ZF\_PHD\_1; 3.  
 DR PROSITE; PS00016; ZF\_PHD\_2; 3.  
 KW DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;  
 FT Transcription regulation; Alternative splicing; Repeat.  
 FT DNA BIND 37 44  
 FT DNA\_BIND 110 117  
 FT DNA\_BIND 357 365  
 FT ZN\_FING 959 1005  
 FT ZN\_FING 1201 1252  
 FT ZN\_FING 1249 1303  
 FT ZN\_FING 1335 1396  
 FT DOMAIN 1449 1471  
 FT DOMAIN 2586 2715  
 FT DOMAIN 26 37  
 FT DOMAIN 248 255  
 FT DOMAIN 362 398  
 FT DOMAIN 402 771  
 FT DOMAIN 808 812  
 FT DOMAIN 1963 1970  
 FT DOMAIN 2251 2259  
 FT DOMAIN

```
FT VARSPLIC 532 582 VSARSRVKTFRFMDPPKPKVEVSPVLRPIITSP
FT VQRPAPVPS -> PLQSLLPMTQLSLSGWAAPTTS
FT ACLDPLSPULLPRCPPLTGLQL (IN ISOFORM
FT TRUNCATED).
FT MISSING (IN ISOFORM TRUNCATED).
FT K -> E (IN REF. 6).
FT CONFLICT 834 834
FT CONFLICT 941 941
FT CONFLICT 1317 1317
FT CONFLICT 1362 1362
FT CONFLICT 1438 1438
FT CONFLICT 1918 1920
FT CONFLICT 2541 2543
FT CONFLICT 2622 2622
FT CONFLICT 2622 2622
SQ SEQUENCE 2715 AA; 293511 MW; C0615B981BBEB7BF CRC64;

Alignment Scores:
Pred. No.: 0.0101 Length: 2715
Score: 155.50 Matches: 78
Percent Similarity: 35.64% Conservative: 20
Best Local Similarity: 28.36% Mismatches: 83
Query Match: 10.67% Indels: 94
DB: 1 Gaps: 15

US-09-931-704-1 (1-797) x TRX2_HUMAN (1-2715)
QY 34 CCACCTCCGCGAGCTCCGGGAGAGCGCCACCGG-----GCC 72
Db 1810 ProSerGluProGlyGlyGluAspProProLeuAspThAspValLeuValPro 1829
QY 73 GGCCCGACCCCGCCCATGGA-----CCTCGAGAGCGGAGCTCGT 114
Db 1830 Gly-AlaProGluArgHisSerProIleGlnAsnLeuAspProProLeuArg- 1846
QY 115 GGGGATGTTAGCTGCTGTCGCGGTGCTGTCGACCTCCCTCGAGTCCAGCTCTCA 174
Db 1847 -----ProAspSerGlySerAlaProProAlaProArgSerPheSe 1861
QY 175 ATCGCACAGG-----GGACCCAGGCGCTGGCCCTCCATCCAGAAA 216
Db 1861 rGlyAlaArgileysValProAsnTySerProSerArgArgProLeuGlyValSe 1881
QY 217 CCTATGACCTCACCGCTACTGGAGCACCAACT-----CCGACGCT 258
Db 1881 rPheGlyProLeuProSerProGlySerProSerSerLeuThrHisIleProThrVa 1901
QY 259 TGGCTGGACCTATCTGAACTACCTGGGCGCCCTTTCAACGAGCAGA-----CT 309
Db 1901 lGly-----AspProAspPheProAlaProProArgArgSerArgProSerProle 1919
QY 310 TCAACCTCCCGCTGGGGCGAGAGCTGCTCCCGAGGCGCACTGTGACTTGGAGGTG 369
Db 1919 uAlaProArgProProProSerArg-----Tr 1928
QY 370 GGCGAAGCTCAATGACAACTGGGGCTGACCCAGAACTACGAGCCCTACAGCCCTTC 429
Db 1928 pAlaSerProLeuIysThrSer-----ProGlnLeuArgValProProThrSe 1946
QY 430 TGTCTTACTTGGTGGCTCAACCTGAGCTGCTGCTGAGCTGCGCCCGCAGCCTGG 489
Db 1946 rValValThrAla-----LeuThrProThrSerGl 1956
QY 490 CCCACTCTGCACCAAGCTCCAGGCGCTGCTGGGCGAGCATTCGGGCGGTGCATGGCAGCTC 549
Db 1956 yGluLeu-----AlaProProGlyProAla----- 1964
QY 550 TGGGCTACCCACTGCCCGCAGCGCTGCTGGGAGTGAACCCACTTGGACTCTGCCCTTG 609
Db 1965 -----ProSerProProPr 1969
QY 610 CCCACAGTACTTCTCCA-----GAAGATGGACGACTTCTGGCTGTGAAGAGCTGC 663
Db 1969 oProGluAspLeuGlyProAspPheGluAspMetGluValValSerGlyLeuSerAlaAl 1989
```

```
QY 664 AGACCTGGCTGTGGCGCTCGGC---CAAGGACTTCAACCGGCTCAAGAGAAGATGCAGC 720
Db 1989 aspLeuaspPheAlaAalaSerLeuLeuGlyThrGluProPheGlnGluGluLeuAla 2009
QY 721 CTCACGACAGC-----TGCACTCACCTCCACCTGGGGGCTCA 757
Db 2009 aAlaGlyAlaMetGlySerSerHisGlyGlyProGlyAspSer 2023

RESULT 5
CA44 BOVIN
ID CA44 BOVIN STANDARD; PRT; 453 AA.
AC Q29442;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 4(IV) chain (fragment).
OS COL4A4.
GN Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 317-328.
RC TISSUE=Lens;
RX MEDLINE=92112769; PubMed=1370461;
RA Mariyama M., Kalluri R., Hudson B.G., Readers S.T.;
RT "The alpha 4(IV) chain of basement membrane collagen. Isolation of
RT cDNAs encoding bovine alpha 4(IV) and comparison with other type IV
RT collagens."
RL J. Biol. Chem. 267:1253-1258(1992).
RN [2]
RP SEQUENCE OF 217-246.
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen."
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 217-233.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowski R.J., Langeveid J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen."
RL J. Biol. Chem. 262:7874-7877(1987).
CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC NIDOGEN.
CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
CC COCHLEA, LUNG AND BRAIN.
CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; M77480; AAA30458.2; ALT SEQ.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001442; ProcollagenC4.  
 DR Pfam; PF01391; Collagen; 4.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagenC4; 2.  
 DR SMART; SM00111; C4; 2.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
 FT NON TER 1  
 FT DOMAIN <1 222 TRIPLE-HELICAL REGION  
 FT DOMAIN 223 453 NONHELICAL REGION (NCU).  
 FT DISULFID 243 332 OR 329 (BY SIMILARITY).  
 FT DISULFID 276 329 OR 332 (BY SIMILARITY).  
 FT DISULFID 288 294 BY SIMILARITY.  
 FT DISULFID 351 449 OR 446 (BY SIMILARITY).  
 FT DISULFID 385 446 OR 449 (BY SIMILARITY).  
 FT DISULFID 397 404 BY SIMILARITY.  
 FT CONFLICT 219 219 I -> P (IN REF. 2 AND 3).  
 SQ SEQUENCE 453 AA; 46384 MW; F7ED410AE9A65BC1 CRC64;

## Alignment Scores:

Pred. No.:	0.0118	Length:	453
Score:	154.50	Matches:	73
Percent Similarity:	38.08%	Conservative:	18
Best Local Similarity:	30.54%	Mismatches:	71
Query Match:	10.60%	Indels:	77
DB:	1	Gaps:	17

US-09-931-704-1 (1-797) x CA44\_BOVIN (1-453)

QY	13	CGGAGCCGCGCTCCCTCCACTCCGACCTCCGCGAGGAGGAG-----CGGCAC	66
DB	18	ProAspProGlyLeu-----ProGlyAspGlnGlyProProGly 30	
QY	67	CGGCGCGCGCCGACCCCGCCGATGAGCTCCGAGC-----	103
DB	31	ProAspGlyProArgGlyValPro-GlyProProGlyProProGlySerValAspLeuLe 50	
QY	104	-AGGGGACTCGTGGGGGAGTTAGCTGCT-----GTGACAGGTGCTCTGGC	150
DB	50	uysGlyGluProGlyAspCysGlyLeuProGlyProProGlyLeuProGlyProProGl 70	
QY	151	ACCTCC-----TGCAGTGCAGCTCTCAATCGCACAGGGGACC	189
DB	70	yProGlyHisGlyPheProGlyCysAsp---GlyHisGlyGlnLysGlyPr 89	
QY	190	CAGGGCTGCGCTCCATCCAGAAACCTATGACTACCCGCTACCTGGAGCACCAC	249
DB	89	oMetGlyPhePro-----GlyProGlnGlyProProGlySerProGl 103	
QY	250	TCCGACGCTGGCTGGGACCTATCTGACTACCTGCGGCCCTCTTCAACGA-----	301
DB	103	yProPro-----GlyAspLysGlyLeuProGlyProProGlyArgGlyProLe 120	
QY	302	-----GCCAGACTTCAACCTCCCGCTGGGGGCGAGAGACTCTCCCGCAGG	348
DB	120	uGlyProProGlySerArgGlyGluProGlyProProAlaAspLeuAspAlaCysProAr 140	
QY	349	CNACTTTGAGTGGAGGTGGCGAAGCTCAATGACAACTGCGGCTGACCCAGAA--	406
DB	140	gileProGlyLeuProGlyValProGlyProArg-----GlyProGluGl 155	
QY	407	-----CTACGAGGCTACAGCACCTTCTGTGTTACTTGGCTCAACCGTC	456
DB	155	yThrMetGlyLeuProGlyMetArgGlyPro-----ProGlyPro-- 168	
QY	457	AGGCTGCCACTGTGAGTGCAGCGCCGAGCCTGGC-----CCACT	495

DB	169	-GlyCys-----LysGlyGluProGlyLeuAspGlyArgGlyGluAspGl 184
QY	496	TTGTCACACGCTCCAGGCGCTGTCGCGGCGCTCATGGCAGCTCTGGGCT 555
DB	184	yLeuProGlySerProGlyProGly---HisLysGlyAspMetGlyGluAlaGlyCy 203
QY	556	ACCCACTGCCCGCGCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCT 608
DB	203	sProGlyAlaProGlyPro-ProGly-----PrometGlyAspProGlyPro 218

## RESULT 6

CA15_HUMAN		
ID	CA15_HUMAN	STANDARD; PRT; 1838 AA.
AC	P20908;	
DT	01-FEB-1991 (Rel. 17, Created)	
DT	01-MAY-1992 (Rel. 22, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Collagen alpha 1(V) chain precursor.	
GN	COL5A1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 556-565.	
RX	MEDLINE=91302336; PubMed=2071595;	
RA	Takahara K., Seto Y., Okasawa K., Okamoto N., Noda A., Yaoi Y.,	
RT	Kato I.;	
RT	"Complete primary structure of human collagen alpha 1 (V) chain.";	
RL	J. Biol. Chem. 266:13124-13129(1991).	
RN	[2]	
RP	SEQUENCE OF 621-822.	
RC	TISSUE=Chorioamniotic membrane;	
RX	MEDLINE=89227189; PubMed=2496661;	
RA	Seyer J.M., Kang A.H.;	
RT	"Covalent structure of collagen: amino acid sequence of three	
RT	cyanogen bromide-derived peptides from human alpha 1(V) collagen	
RT	chain.";	
RL	Arch. Biochem. Biophys. 271:120-129(1989).	
RN	[3]	
RP	SEQUENCE OF 823-950, AND HEPARIN-BINDING.	
RX	MEDLINE=90366601; PubMed=2203476;	
RA	Yaoi Y., Hashimoto K., Koitabashi H., Takahara K., Ito M., Kato I.;	
RT	"Primary structure of the heparin-binding site of type V collagen.";	
RL	Biochim. Biophys. Acta 1035:139-145(1990).	
RN	[4]	
RP	SEQUENCE OF 556-571.	
RC	TISSUE=Placenta;	
RX	MEDLINE=92239022; PubMed=1571108;	
RA	Mann K.;	
RT	"Isolation of the alpha 3-chain of human type V collagen and	
RT	characterization by partial sequencing.";	
RL	Biol. Chem. Hoppe-Seyler 373:69-75(1992).	
RN	[5]	
RP	SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.	
RC	TISSUE=Chorioamniotic membrane;	
RX	MEDLINE=94237164; PubMed=8181482;	
RA	Moradi-Ameli M., Rousseau J.C., Klemen J.P., Champliand M.F.,	
RA	Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;	
RT	"Diversity in the processing events at the N-terminus of type-V	
RT	collagen.";	
RL	Eur. J. Biochem. 221:987-995(1994).	
RN	[6]	
RP	DISEASE, AND VARIANT EDS-I SER-1639.	
RX	MEDLINE=97195540; PubMed=9042913;	
RA	de Paeppe A., Nuytinck L., Hauser I., Anton-Lamprecht I.,	
RA	Naeyaert J.-M.;	
RT	"Mutations in the COL5A1 gene are causal in the Ehlers-Danlos	
RT	syndromes I and II.";	
RL	Am. J. Hum. Genet. 60:547-554(1997).	
CC	-!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN	
CC	(FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE	





DR SMART; SM00549; TAFH; 1.  
 KW Transcription regulation; Nuclear protein.  
 FT DOMAIN 39 42 POLY-HIS.  
 FT DOMAIN 52 57 POLY-ALA.  
 FT DOMAIN 98 101 POLY-GLY.  
 FT DOMAIN 142 148 POLY-ALA.  
 FT DOMAIN 268 275 POLY-PRO.  
 FT DOMAIN 331 337 POLY-ALA.  
 FT DOMAIN 680 683 POLY-PRO.  
 FT DOMAIN 808 813 POLY-ALA.  
 FT DOMAIN 828 831 POLY-ASP.  
 FT CONFLICT 105 117 PGPSPPRRPLVPA -> GRGLLQRRGGRES (IN REF. 3).  
 FT CONFLICT 136 136 A -> S (IN REF. 2).  
 FT CONFLICT 185 185 G -> GPG (IN REF. 2).  
 FT CONFLICT 233 264 MISSING (IN REF. 3).  
 FT CONFLICT 293 293 P -> L (IN REF. 3).  
 SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;

Alignment Scores:  
 Pred. No.: 0.0126 Length: 1083  
 Score: 154.00 Matches: 86  
 Percent Similarity: 35.76% Conservative: 17  
 Best Local Similarity: 29.86% Mismatches: 108  
 Query Match: 10.56% Indels: 77  
 DB: 1 Gaps: 16

US-09-931-704-1 (1-797) x T2D3\_HUMAN (1-1083)

QY 1 ATTAAGCTTCGCGGAGCCGCGG-----CTCGCC 30  
 Db 146 ValAlaAlaGlyProGluProAlaProAlaGlyProAlaLysProAlaGlyProAlaAla 165  
 QY 31 CTCCACTCGCCAGCCTCGGAGAGAGCGGCACCGCGCGGCGCCAGCCGCCAGCCCA 90  
 Db 166 LeuAlaAlaArgAlaGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 184  
 QY 91 TGGACCTCCG-----AGCAGGGGACTCGTGGGGGATGTTAGC---GTGCC 132  
 Db 185 -GlyLysProAlaGlyProGlyAlaAlaGlnThrLeuAsnGlySerAlaAlaLeuLeuAs 204  
 QY 133 TGTGCAGGTGCTTGGCAGCTCCCTGCGAGTGCCAGTCTCAATCGCACAGGGGAGCCAG 192  
 Db 204 nSerHisAlaAlaPro-----AlaValSerLeuValAsnAsnGlyProAl 221  
 QY 193 GGCCTGCGCCCTCCATCCAGAAACCTATGACCTACCGCTACTGAGCAGCCACTCC 252  
 Db 221 aAla-----LeuLeuProLeuProLysProAlaAlaProGly-- 233  
 QY 253 GCAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCGCCCTTTCAACGAGCCAGACTTCA 312  
 Db 234 -----ThrValIleGlnThrProProPheValGlyAlaAlaAlaPr 247  
 QY 313 ACCCTCCCG-----CCTGGGCGAGAGACTCTGCCAGGGCCACTGTTGACT 360  
 Db 247 oProAlaProAlaProSerProAlaAlaProAlaProAlaProAlaAlaAla 267  
 QY 361 TGGAGGTGTGGGAGCCTCAATGCAACTCGCGCTGACCCAGAACTACGAGGCTTACA 420  
 Db 267 aProProProProProProAlaProAlaAlaThrLeuAlaArgProProGlyHisProAlaGl 287  
 QY 421 GCCACTTCTGTGTACTTGGTGCCCTCAACCG-----TCAGGCTGCCACTG 468  
 Db 287 yProProThraAlaAlaProAlaValProProProAlaAlaAlaGlnAsnGlyGlySerAl 307  
 QY 469 CTGAGTGTGGCGGAGCCTGCGCCACTTCTGCACACGAGCTCCAGGCGCTGCTGGGAGCA 528  
 Db 307 aGlyAlaAlaProAlaProAlaProAlaAlaGlyGlyProAlaGlyValSerGlyGln-- 326  
 QY 529 TTGGGGGGCTATGCGAGCTCTGGGCTACCCACTGCGCCAGCGCGCTGCTGGGACTGAAC 588  
 Db 327 ---ProGlyAlaAlaAlaAlaAla----- 335

QY 589 CCACCTTGGACTCTCGCGCCCTGCCCA-----CA 615  
 Db 336 -----ProAlaProGlyValLysAlaGluSerProLysArgValValGl 350  
 QY 616 GTGACTTCTCCAGAGATGGACGACTTCTGGCTGTGAGGAGTGCAGACCTGGCT-- 673  
 Db 350 nAlaAlaProProAla---AlaGlnThrLeuAlaAlaSerGlyProAlaSerThrAlaAl 369  
 QY 674 ----GTGGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAAGATGCACGCTCCACAG 729  
 Db 369 aSerMetValIleGlyProThrMetGlnGlyAlaLeuProSerProAlaAla----- 386  
 QY 730 CTCGAGTCACCTCGCACCTGGG 751  
 Db 387 -ValProProProAlaProGly 393

RESULT 8  
 CA35\_HUMAN STANDARD; PRT; 382 AA.  
 ID CA35\_HUMAN  
 AC P25940;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 3(V) chain (fragments).  
 GN COL5A3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Placenta;  
 RX MEDLINE=92239022; PubMed=1571108;  
 RA Mann K.;  
 RT "Isolation of the alpha 3-chain of human type V collagen and  
 RT characterization by partial sequencing."  
 RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).  
 CC -!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 CC (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE  
 CC COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS  
 CC TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.  
 CC -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN  
 CC MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND  
 CC ONE ALPHA 3(V) CHAINS IN PLACENTA.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -!- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 1(XI) CHAINS.  
 DR PIR; S20375; S20375.  
 DR MIM; 120216; -.  
 DR InterPro; IPR000087; Collagen.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Collagen.  
 FT NON\_TER 1  
 FT NON\_CONS 86 87  
 FT NON\_CONS 131 132  
 FT NON\_CONS 167 168  
 FT NON\_CONS 200 201  
 FT NON\_CONS 284 285  
 FT NON\_CONS 317 318  
 FT NON\_CONS 359 360  
 FT NON\_TER 382 382  
 SQ SEQUENCE 382 AA; 35352 MW; F80BB5968F696B5C CRC64;

Alignment Scores:  
 Pred. No.: 0.0169 Length: 382  
 Score: 152.00 Matches: 78  
 Percent Similarity: 28.84% Conservative: 14  
 Best Local Similarity: 24.45% Mismatches: 74  
 Query Match: 10.43% Indels: 153  
 DB: 1 Gaps: 15

US-09-931-704-1 (1-797) x CA35\_HUMAN (1-382)



Db 171 AlaSerGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGly 190  
QY 507 GGCTGG-----TGAGAGTGGGCGAGCTGGGCGGAGCTCAG----- 469  
Db 191 GlyTyrGlyProGlyGlnGlnGlyProSerGlyProGlySerAlaAlaAlaAlaAla 210  
QY 468 -----CAGTGGCAGCTCAGCGTTGAGGCGCAGCGCAAGTAACACAGAA 427  
Db 211 AlaAlaSerGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyPro 230  
QY 426 GTGGCTGTAGGCT-----CGTAGTCTGGGTGAGCGGCGGAGCTCTGCTGCTCAT 382  
Db 231 GlyGlyTyrGlyProGlyGlnGlnGlyLeuSerGlyProGlySerAlaAlaAlaAla 248  
QY 381 TGAGGCTTCCGACACCTCCAGTCCAGTGGGCGGCTGGGCGAGCTCTGCTGCTGCTGCT 322  
Db 249 ----AlaAlaAlaAlaGlyProGlyGlnGlnGlyProGlyGlyTyrGly-----ProGly 265  
QY 321 GGGAGGCTTGAAGTCTGGCTCGTTGAAAGGGGGCGGCGGAGTTCAGATAGTCCAG 262  
Db 266 GlnGlnGlyProSer-----GlyProGlySerAlaAlaAlaAlaAla 278  
QY 261 CCAAGCTCGGAGTGTGTCTCCAGTGGGCGGAGTTCAGATAGTTCGTGATGGAGG 202  
Db 279 AlaAlaAlaAlaAlaGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyTyr 298  
QY 201 GGCAGGCC---CTGGGTCCCTCTGCGATTGAGAGCTGGGAGTGGGAGTGGCAGA 145  
Db 299 GlyProGlyGlnGlnGlyPro-----SerGlyAlaGly 309  
QY 144 GCACCGTGCAGCAGCGCTAAACATCCGCCAGAGTCCCTGCTGGAGGTCCTAGTGGGCG 85  
Db 310 SerAlaAlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyLeuGlyGlyTyrGlyPr 329  
QY 84 TGGG---GCTGGGCGGCGGCGGCTGGGCGG-----TCCTCTCCGAGGCTGGGCGAG 37  
Db 329 oGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyTyrGlyPr 349  
QY 36 TGGGAGGCGGAGCGGCGCTCCGCGAGCT 6  
Db 349 oGlySerAlaSerAlaAlaAlaAlaAla 359  
RESULT 10  
PRP3 MOUSE STANDARD; PRT; 296 AA.  
AC P05143;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE Proline-rich protein MP-3 (Fragment).  
GN PRP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86059475; PubMed=2999141;  
RA Ann D.K.; Carlson D.M.;  
RT "The structure and organization of a proline-rich protein gene of a  
RT mouse multigene family."  
RL J. Biol. Chem. 260:15863-15872 (1985).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M12100; AAA40005.1; -.  
DR MGD; MGI:97773; Prp.

KW Repeat; Saliva.  
FT NON TER 1  
SQ SEQUENCE 296 AA; 29521 MW; 7F146824E8AF3269 CRC64;  
Alignment Scores:  
Pred. NO.: 0.0279 Length: 296  
Score: 148.50 Matches: 82  
Percent Similarity: 36.92% Conservative: 14  
Best Local Similarity: 21.54% Mismatches: 87  
Query Match: 10.19% Indels: 77  
DB: 1 Gaps: 18  
US-09-931-704-1 (1-797) x PRP3\_MOUSE (1-296)  
QY 12 GCCGAGCGGCGCTGCGCTCCACT-----CGGCGAGCTCCGGG 53  
Db 16 SerGlySerGlnProArgProValAenGlySerGlnGlnGlyProPro-ProProGl 35  
QY 54 A-----GAGGAGCGCAGCCCGCGCGC-----CGAGCCCC 83  
Db 35 yGlyProGlnProArgProProGlnGlyProProProGlyGlyProGlnProArgPr 55  
QY 84 AGCCCATGAGCTCCGACAGGGGAGTCTGGGGGATGTTAGCGTCCCTGTGACCGGT 143  
Db 55 o-ProGlnGlyProProProGly-----GlyProGlnProArgProGlnGlyP 73  
QY 144 CTCTGGCAGCTCCCTGCGAGTCCAGCTCTCAATCGCACAGGGGAGCCAGGCGCTGGCCCC 203  
Db 73 to---ProProProGlyGlyProGlnProArgProProGlnGlyProProPro---ProG 91  
QY 204 TCCATCCAGAAAACCTATGACCTCACCGCTACCTCGAGCAGCCCAACTCCCGAGCTTGGCT 263  
Db 91 lyGlyProGln-----ProArgPro---ProGlnGlyProProPro----- 103  
QY 264 GGGAGCTATCTGAATACCTGGGCGGCGCTTTCAACAGCAGCAGACTTCAACCTCCCGCG 323  
Db 104 -----ProGlyGlyProGlnGlnArgProProGlnGlyProProProp 118  
QY 324 CTGGGGG-----AGAGACTCTGCCGAGGCGCACTGTTGACTTGGAGGTGGCGAAGC 377  
Db 118 roGlyGlyProGlnGlnArgProProGlnGlyProProProGlyGly----- 134  
QY 378 CTCAATGACAAACTCGGCTGACCCAGAACTACGAGGCGCTACGAGCCACCTTCTGTGTAC 437  
Db 135 -----ProGlnProArgProProGlnGlyPro----- 143  
QY 438 TTGCGTGGCTCAACCGTCAAGCTGCGCTGCGCTGCTGAGCTGCGCGCAGCTTGGCCACTTC 497  
Db 144 -----ProProProAlaGly-----ProGlnProArgPro---P 154  
QY 498 TGCACAGCTCCAGGCGCTGGGCGAGCATTC-----GGGCGTCATGGCA 545  
Db 154 roGlnGlyProProProProAlaGlyProHisLeuArgProThrGlnGlyProProPro 174  
QY 546 GCTCTGGG-----CTACCCACTGCGCCCGCGCTGCTCTGCTGCTGCTGCTGCTGCTGCT 590  
Db 174 hrGlyGlyProGlnGlnArgTyrProGlnSerPro-ProProGlyGlyProGlnPro 193  
QY 591 ACTTGGAGCTCTGGCGCTGCCACAGTACTTCTCCAGAAAGTGGACGACTTCTGCTGCTG 650  
Db 194 ArgProProGlnGlyPro-ProProProGlyGlyProHisProArgProThrGlnGly-- 212  
QY 651 CTGAGGAGCTGCAGACCTGGCTGGGCGCTGGGCGCTGGGCGCTTCAACCGCG 702  
Db 213 -----ProProProThrGlyProGlnProArgProThrGlnGly 225  
RESULT 11  
ID CA34 HUMAN STANDARD; PRT; 1670 AA.  
CA34 HUMAN  
AC Q01955; Q9BQ12;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)



DE Collagen alpha 3 (IV) chain precursor (Goodpasture antigen).  
GN COL4A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Kidney;  
RC MEDLINE=94364994; PubMed=8083201;  
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;  
RT "Complete primary structure of the human alpha 3(IV) collagen chains in  
human tissues.";  
RL J. Biol. Chem. 269:23013-23017(1994).  
[2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Kidney;  
RC MEDLINE=94364994; PubMed=8083201;  
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;  
RT "Complete primary structure of the human alpha 3(IV) collagen chains in  
human tissues.";  
RL J. Biol. Chem. 269:23013-23017(1994).  
[3]  
RN REVISIONS  
RA Leinonen A.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
[4]  
RN SEQUENCE FROM N.A., VARIANTS AS E-297; R-407; R-640; R-1167; E-1207;  
RP Q-1215; S-1277; T-1330; E-1334; E-1347 AND C-1661, AND VARIANTS R-43;  
RP E-162; Y-326; H-408; R-451; L-574; E-1269 AND P-1474.  
RX MEDLINE=21064696; PubMed=1134255;  
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,  
RT Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
autosomal Alport syndrome.";  
RL J. Am. Soc. Nephrol. 12:97-106(2001).  
[5]  
RN SEQUENCE OF 1453-1670 FROM N.A.  
RP MEDLINE=91353570; PubMed=1882840;  
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reiders S.T.;  
RT "Sequence and localization of a partial cDNA encoding the human alpha  
3 chain of type IV collagen.";  
RL Am. J. Hum. Genet. 49:545-554(1991).  
[6]  
RN SEQUENCE OF 1331-1670 FROM N.A.  
RP TISSUE=Kidney;  
RC MEDLINE=92147878; PubMed=1737849;  
RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
RT Pusey C.D.;  
RT "Molecular cloning of the human Goodpasture antigen demonstrates it  
to be the alpha 3 chain of type IV collagen.";  
RL J. Clin. Invest. 89:592-601(1992).  
[7]  
RN SEQUENCE OF 1644-1670 FROM N.A.  
RP TISSUE=Kidney;  
RC MEDLINE=94364994; PubMed=8083201;  
RA Ding J.;  
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
[8]  
RN SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.  
RC TISSUE=Kidney;  
RX MEDLINE=94124597; PubMed=8294492;  
RA Feng L., Xia Y., Wilson C.B.;  
RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)  
collagen gene. Differential expression of mRNA transcripts that  
predict three protein variants with distinct carboxyl regions.";  
RL J. Biol. Chem. 269:2342-2348(1994).  
[9]  
RN SEQUENCE OF 1-29 FROM N.A.  
RP MEDLINE=98196854; PubMed=9537506;  
RA Monota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,  
RA Ninomiya Y.;  
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and

alpha4(IV) collagen chains are arranged head-to-head on chromosome  
2q36.";  
RL FEBS Lett. 424:11-16(1998).  
[10]  
RN ALTERNATIVE SPLICING.  
RP MEDLINE=93280184; PubMed=8505332;  
RA Bernal D., Quinones S., Saus J.;  
RT "The human mRNA encoding the Goodpasture antigen is alternatively  
spliced.";  
RL J. Biol. Chem. 268:12090-12094(1993).  
[11]  
RN VARIANT PRO-1474.  
RP MEDLINE=95078827; PubMed=7987301;  
RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,  
RA Reiders S.T., Smeets H.J.M.;  
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
recessive Alport syndrome.";  
RL Hum. Mol. Genet. 3:1269-1273(1994).  
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE',  
MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/  
NIDOCEN.  
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS. ALPHA 1(IV)-  
ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2/V AND  
3/L5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR  
C-TERMINAL NC1 DOMAINS.  
CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE  
COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,  
COCHLEA, LUNG AND BRAIN.  
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
DOMAIN (NC1) AT THEIR C-TERMINUS. FREQUENT INTERRUPTIONS OF THE  
G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
TRIPLE-HELICAL 7S DOMAIN.  
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -1- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL  
N-LINKED GLYCOSYLATION SITE.  
CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
IV COLLAGENS.  
CC -1- PTM: Phosphorylated by the Goodpasture antigen-binding protein.  
CC -1- DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE  
THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS  
CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.  
CC -1- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I  
AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY  
GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,  
HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN  
MALES AND FEMALES.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; X80031; CAA56335.1; --  
DR EMBL; AJ288487; CAC36101.1; --  
DR EMBL; AJ288488; CAC36101.1; JOINED.  
DR EMBL; AJ288489; CAC36101.1; JOINED.  
DR EMBL; AJ288490; CAC36101.1; JOINED.  
DR EMBL; AJ288491; CAC36101.1; JOINED.  
DR EMBL; AJ288492; CAC36101.1; JOINED.  
DR EMBL; AJ288493; CAC36101.1; JOINED.  
DR EMBL; AJ288494; CAC36101.1; JOINED.

DR EMBL: AJ288495: CAC36101.1: JOINED.  
 DR EMBL: AJ288496: CAC36101.1: JOINED.  
 DR EMBL: AJ288497: CAC36101.1: JOINED.  
 DR EMBL: AJ288498: CAC36101.1: JOINED.  
 DR EMBL: AJ288499: CAC36101.1: JOINED.  
 DR EMBL: AJ288500: CAC36101.1: JOINED.  
 DR EMBL: AJ288501: CAC36101.1: JOINED.  
 DR EMBL: AJ288502: CAC36101.1: JOINED.  
 DR EMBL: AJ288503: CAC36101.1: JOINED.  
 DR EMBL: AJ288504: CAC36101.1: JOINED.  
 DR EMBL: AJ288505: CAC36101.1: JOINED.  
 DR EMBL: AJ288506: CAC36101.1: JOINED.  
 DR EMBL: AJ288507: CAC36101.1: JOINED.  
 DR EMBL: AJ288508: CAC36101.1: JOINED.  
 DR EMBL: AJ288509: CAC36101.1: JOINED.  
 DR EMBL: AJ288510: CAC36101.1: JOINED.  
 DR EMBL: AJ288511: CAC36101.1: JOINED.  
 DR EMBL: AJ288512: CAC36101.1: JOINED.  
 DR EMBL: AJ288513: CAC36101.1: JOINED.  
 DR EMBL: AJ288514: CAC36101.1: JOINED.  
 DR EMBL: AJ288515: CAC36101.1: JOINED.  
 DR EMBL: AJ288516: CAC36101.1: JOINED.  
 DR EMBL: AJ288517: CAC36101.1: JOINED.  
 DR EMBL: AJ288518: CAC36101.1: JOINED.  
 DR EMBL: AJ288519: CAC36101.1: JOINED.  
 DR EMBL: AJ288520: CAC36101.1: JOINED.  
 DR EMBL: AJ288521: CAC36101.1: JOINED.  
 DR EMBL: AJ288522: CAC36101.1: JOINED.  
 DR EMBL: AJ288523: CAC36101.1: JOINED.  
 DR EMBL: AJ288524: CAC36101.1: JOINED.  
 DR EMBL: AJ288525: CAC36101.1: JOINED.  
 DR EMBL: AJ288526: CAC36101.1: JOINED.  
 DR EMBL: AJ288527: CAC36101.1: JOINED.  
 DR EMBL: AJ288528: CAC36101.1: JOINED.  
 DR EMBL: AJ288529: CAC36101.1: JOINED.  
 DR EMBL: AJ288530: CAC36101.1: JOINED.  
 DR EMBL: AJ288531: CAC36101.1: JOINED.  
 DR EMBL: AJ288532: CAC36101.1: JOINED.  
 DR EMBL: AJ288533: CAC36101.1: JOINED.  
 DR EMBL: AJ288534: CAC36101.1: JOINED.  
 DR EMBL: AJ288535: CAC36101.1: JOINED.  
 DR EMBL: AJ288536: CAC36101.1: JOINED.  
 DR EMBL: AJ288537: CAC36101.1: JOINED.  
 DR EMBL: AJ288538: CAC36101.1: JOINED.  
 DR EMBL: M92993: AAA21610.1: -.  
 DR EMBL: M5790: AAB19637.1: -.  
 DR EMBL: M81379: AAA51556.1: -.  
 DR EMBL: L08650: AAA52044.1: -.  
 DR EMBL: U02519: AAA18942.1: -.

## Alignment Scores:

Pred. No.: 0.0275  
 Score: 148.50  
 Percent Similarity: 35.00%  
 Best Local Similarity: 29.23%  
 Query Match: 10.19%  
 DB: 1

Length: 1670

Matches: 76

Conservative: 15

Mismatch: 97

Indels: 18

Gaps: 18

US-09-931-704-1 (1-797) x CA34\_HUMAN (1-1670)

QY 1 ATTAAGCTTCGCGGAGCGGCTCGCCCTCCACATC-----CGC 42  
 Db 576 VallysglyLeuProGlyProGlyGlyGluLeuAlaLeuSerGlyGlyGlyAspGln 595  
 QY 43 CAGCTCCGGGAGAG-----GAGCGCACCAGCGCGCGCCAGCC----- 81  
 Db 596 GlyProGlyAspProGlySerProGlyProAlaGlyProAlaGlyPro 615  
 QY 82 CCAGCCCCATGCTCCGAGCAGG---GGACTCTGGGGGATGTTAGCGTCCCTGTGCA 138  
 Db 616 ProGly-TyrGlyProGlnGlyGluProGlyLeuGlnGlyThrGlnGlyValPro----- 633  
 QY 139 CGGTGCTCTGGCACCTCCCTCAGTGTCCAGTCTCTCAATCGCACAGGGGAGCCAGGCGCTG 198

Db 634 -GlyAlaProGlyProGlyGluAla-----GlyProArgGly-- 646  
 QY 199 GCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTGGAGCACCACCTCCGAGCT 258  
 Db 647 -----GluLeuSerValSerThrProValProGlyProProGlyPro----- 660  
 QY 259 TGGCTGGGACCTATCTGAACCTACCTGGGCGCCCTTTCAACGAGCAGCAGCTTCAACCCCTC 318  
 Db 661 -----ProGlyProPro-----GlyHisProGlyProG 670  
 QY 319 CCCGCTGGGGCAGAGACTCTGCCAGGCGCAGCTTGTGAGTGGAGGTGTGCGGAGCC 378  
 Db 670 nGlyProGlyIleProGlySerLeuGlyGlyCysGlyAspProGlyLeuProGlyPr 690  
 QY 379 TCAATGACAAACTGCGGCTGACCCAGACTACGAGCGCTTACAGCCACCTCTCTGTGTACT 438  
 Db 690 oAsp-----GlyGluProGlyIleProGlyIleGlyPheProGlyPro----- 704  
 QY 439 TCGGTGGCTCAACCGTCA-----CAGCTGTC-----GGCTG 462  
 Db 705 -----ProGlyProGlyGlyAspGlnGlyPheProGlyThrLysGlySerLeuGlyCy 722  
 QY 463 COACTGCTGAGCTGCGCG-----CAGCTGTC----- 490  
 Db 722 s-----ProGlyLysMetGlyGluProGlyLeuProGlyLysProGlyLe 737  
 QY 491 -----CCACTTCTGCACCGCTCCAGCGCGCTGCT----- 520  
 Db 737 uProGlyAlaLysGlyGluProAlaValAlaMetProGlyGlyProGlyThrProGlyPh 757  
 QY 521 ---GGGCGAG-----CATTGGCGGCGCTCATGCGAGCTCTGGGCTACCCACTGCCCGCAGC 570  
 Db 757 eProGlyGluArgGlyAsnSerGlyGlyGluIleGlyLeuProGlyLeuProGly 777  
 QY 571 CGCTCCCTGGGACTGAA-----CCACTTGGACTCTCTGGC 605  
 Db 777 y-LeuProGlyThrProGlyAsnGluGlyLeuAspGlyProArgGlyAspProGly 795  
 RESULT 12  
 ID PRPM\_HUMAN STANDARD; PRT; 234 AA.  
 AC P10161; P02813;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Salivary proline-rich protein PO (Allele M) [Contains: Peptide P-D]  
 DE (Fragment).  
 GN PRB4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=89121439; PubMed=3220251;  
 RA Lyons K.M., Stein J.H., Smithies O.;  
 RT "Many protein products from a few loci: assignment of human salivary  
 RT proline-rich proteins to specific loci."  
 RL Genetics 120:255-265 (1988).  
 RN [2]  
 RP MEDLINE=8186122; PubMed=6841349;  
 RA Saitoh E., Isemura S., Sanada K.;  
 RT "Complete amino acid sequence of a basic proline-rich peptide, P-D,  
 RT from human parotid saliva."  
 RL J. Biochem. 93:495-502 (1983).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR PIR; A03295; PIHUSD.  
DR PIR; S03175; S03175.  
DR MIM; 168730; -;  
DR MIM; 180990; -;  
KW Repeat; Parotid gland; Saliva; Multigene family.  
FT NON TER 1 234 PEPTIDE P-D.  
FT CHAIN 165 234  
SQ SEQUENCE 234 AA; 23676 MW; 310AFPI3A44E747F CRC64;

Alignment Scores:  
Pred. No.: 0.0322 Length: 234  
Score: 147.50 Matches: 75  
Percent Similarity: 31.62% Conservative: 11  
Best Local Similarity: 27.57% Mismatches: 72  
Query Match: 10.12% Indels: 114  
DB: 1 Gaps: 17

US-09-931-704-1 (1-797) x PRPM\_HUMAN (1-234)

Qy 13 CGAGGCGGGCTGCTCCCTCCACTCCGCGAGCTCCG----- 51  
Db 37 ProProHisProGlyLysProGluArgProProGlnGlyAsnGlnSerGln 56  
Qy 52 GGAGGAGGCGCACCGCGCC----- 72  
Db 57 GlyProProHisProGlyLysProGluArgProProGlnGlyAsnGlnSer 76  
Qy 73 ---GGCCGAGCC-----CAGCCCATGACCTCCGAGCAGGGGACTCGTGG 117  
Db 77 GlnGlyPro-ProProThrProGlyLysProGluGlyProProGlnGlyAsnGln 96  
Qy 118 GGATTTAGCTGCTGTGACGGTCTCTG-----GCACCTCCCTGCGAGTCCAGCTC 171  
Db 96 nSerGlnGlyProProHisProGlyLysProGluArgProProGlnGlyAsn 116  
Qy 172 TCAATCGCAGGCGGACCCAGGCGCTGGCCCTCCATCCAGAAACCTATGACCTCACCC 231  
Db 116 nSerHisArg-----ProProPr 123  
Qy 232 GCTACTCGAGCACCA-----ACTCCGAGCTTGGTGGAGCTTGAATACCT-- 283  
Db 123 oProProGlyLysProGluArgProProGlnGlyAsnGlnSerGlnGlyProPr 143  
Qy 284 -----GGGCCCCCTTTCAA-----CGAGCCAGACTTCA 312  
Db 143 oProHisProGlyLysProGluGlyProProGlnGlyAsnLysSerArgSerAl 163  
Qy 313 ACCCTCCCGCTGGGGGAGAGACTCTGCCAGGCGCACTGTGACTTGGAGGTGTGGC 372  
Db 163 aArgSerProGly----- 168  
Qy 373 GAAGCTCAATGACAACTCGGCTGACCCAGACTACAGGCTTACAGCCACTTCTGT 432  
Db 169 -LysProGln-----GlyProGlnGlnGlyAsnLysProGlnGly-- 183  
Qy 433 GTTACTTGGCTGGCTCAACCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 492  
Db 184 -----ProProProGly-----LysProGlnGlyProPr 194  
Qy 493 ACTTCTGCACCACTTCCAGGCGCTGCTGGGAGCATTCGGGCGTCATGCGCTCTGG 552  
Db 194 o-----ProProGly-----GlyAsnProGl 201  
Qy 553 GCTACCACTGCCCCAGCGCTCTCTCTGGAGTGAACCCACTTGAATCTCTGCGCTGCCC 612  
Db 201 nGlnProGlnAlaProProAlaGly-----LysProGlnGlyProProPr 216  
Qy 613 ACAGTGACTTCTTCCAGAGATGGACGACTTCTGGCTGTGAAGAGCTGCAGACTGGC 672

Db 216 o-----ProProGlnGlyArg-----Pr 223  
Qy 673 TGTGCGCTCGCCCAAGGACTTCAACCGGCTCAA 706  
Db 223 oProArgProAlaGlnGlyGlnProProGln 234

RESULT 13  
YQ35 CAEEL  
ID YQ35 CAEEL STANDARD; PRT; 317 AA.

AC Q09456;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Putative cuticle collagen C09G5.5.  
GN C09G5.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OK NCBI\_TaxID=6239;  
RN [1]

SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Palmer S.;

RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.

CC -! FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).

CC -! SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-

CC LINKS (BY SIMILARITY).

CC -! SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z46791; CAA86758.1; -;

DR WormPep; C09G5.5; C01485.

DR InterPro; IPR002486; Col.cuticle\_N.

DR Pfam; PF01391; Collagen; 3.

DR Pfam; PF01484; Col.cuticle\_N; 1.

KW Hypothetical protein; Cuticle; Connective tissue; Repeat;

KW Multigene family; Collagen.

FT DOMAIN 92 124 TRIPLE-HELICAL REGION.

FT DOMAIN 137 199 TRIPLE-HELICAL REGION.

FT DOMAIN 202 264 TRIPLE-HELICAL REGION.

SQ SEQUENCE 317 AA; 31283 MW; 685DCP24612707BB CRC64;

Alignment Scores:

Pred. No.: 0.0321 Length: 317  
Score: 147.50 Matches: 68  
Percent Similarity: 36.97% Conservative: 10  
Best Local Similarity: 32.23% Mismatches: 95  
Query Match: 10.12% Indels: 38  
DB: 1 Gaps: 10

US-09-931-704-1 (1-797) x YQ35 CAEEL (1-317)

Qy 24 GCTCGCTCCCTCCACTCCGCA---GCCTCCGGGAGGAGCGCACCCGCGGCCAGC 80

Db 91 AlaGlyProProGlyProGlyAlaSerGlyAspArgGlyLeuAspGlyGlnProGly 110

Qy 81 CCCAGCCCCATGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTGCTGTGCAG 140

Db 111 ProAlaGlyLys--ProGlyGlnProGlyValAlaGly-----ProAlaHis 126

Qy 141 GTGCTCTGGCACCTCCCTGCAGTCCAGCTCTCAATCGCAGGAGGACCCAGGCTGCG 200

Db 126 isGln--GlnGlnGluCysileLysCysProGlnGlyAlaProGlyProAlaGlyAlaP 145  
 QY 201 CCTCCATCCAGAAAACCTATGACTCACCCTGCTACCTGGAGCACCACCACTCGCAGCTTG 260  
 Db 145 roGlyAsnProGlyProGlnGlyProGlnGlyAsnProGlyAlaProAlaHisGlyGly 165  
 QY 261 GCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCC 320  
 Db 165 ly-----GlnGlyPro-----ProGlyP 171  
 QY 321 CGCTGGGGGAGAGACTCTCCAGGCGCCACTGTTGACTGAGGTGGGAAGCCTC 380  
 Db 171 roProGlyProAlaGlyAspAlaGlySerProGlyGlnAlaGlyAlaProGlyAsnProG 191  
 QY 381 AATGACAACTGCGGCTGACCCAGAACTACGAGGCTCAGCCACTCTGTGTTACTTG 440  
 Db 191 lyArgProGlyGlnSerGlyGlnArgSerArgGlyLeuProGlyProSer-----G 208  
 QY 441 CGTGGCTCAACCTCAGGCTGCCACTGCTGAGCTGCGCGC----- 481  
 Db 208 lyArgProGlyProGlnGlyProPro-----GlyAlaProGlyGlnProGlySerGlyS 226  
 QY 482 --CAGCTGGCGCCACTT---CTGACACGAGCTCCAGGCGCTGCTGGGAG-----CA 530  
 Db 226 erThrProGlyProAlaGlyProProGlyProGlyProGlyProGlyProGlyHisP 246  
 QY 531 GCGGGCTCATGGCAGCTCTGGGCTACCCACTGCCAGCCGCTGCCCTGGGAGCTGAACCC 590  
 Db 246 roGlyGlnAspGlyGlnProGlyAlaProGlyAsnAspGlyAla-ProGlySerAspAla 265  
 QY 591 ACTTGGACTCTGGCGCTGCCACAGT 617  
 Db 266 AlaTyrCysProCysProAlaArgSer 274

## RESULT 14

GP1\_CHLRE STANDARD; PRT; 555 AA.  
 AC Q9FPQ6; Q03927;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich  
 DE Glycoprotein 1).  
 GN GP1.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21159092; PubMed=11258910;  
 RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,  
 RA Goodenough U.W.;  
 RA "Glycosylated polyproline II rods-with-kinks as a structural motif in  
 RT plant hydroxyproline-rich glycoproteins.";  
 RL Biochemistry 40:2978-2987(2001).  
 [2]  
 RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.  
 RX MEDLINE=91017504; PubMed=1699225;  
 RA Adair W.S., Apt K.E.;  
 RA "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs  
 RT encoding cell wall hydroxyproline-rich glycoproteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).  
 CC -!- FUNCTION: Major component of the outer cell wall w6 (crystalline)  
 CC layer.  
 CC -!- SUBUNIT: Associates with GP2 and GP3.  
 CC -!- PTM: N-glycosylated and O-glycosylated.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF309494; AAG45420.1; -;  
 DR EMBL; M58496; AAG49706.1; ALT\_SEQ.  
 DR GlycoSuiteDB; Q9FPQ6; -;  
 DR InterPro; IPR002965; P rich extensin.  
 DR PRINTS; PR01217; PRICHEXTENSIN.  
 DR PRINTS; PR01218; PSILEXTENSIN.  
 KW Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 29  
 FT CHAIN 30 555  
 FT DOMAIN 40 339  
 FT DOMAIN 259 279  
 FT CARBOHYD 399 399  
 FT CARBOHYD 455 455  
 FT CARBOHYD 493 493  
 SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.032 Length: 555  
 Score: 147.50 Matches: 65  
 Percent Similarity: 33.04% Conservative: 9  
 Best Local Similarity: 29.02% Mismatches: 103  
 Query Match: 10.12% Indels: 47  
 DB: 1 Gaps: 7

US-09-931-704-1 (1-797) x GP1\_CHLRE (1-555)

QY 1 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCGCCAGCTCCGGGAGAGGAG 60  
 Db 167 ValProSerProAlaProSerProThrProSerProSerProSerProValPro 186  
 QY 61 CCGCACCAGCGCGCGCCAGCCAGCCCATGAGCTCCGAGCAGGGGACTCGTGGGGGA 120  
 Db 187 ProSerProAlaProSerProAlaPro-ProValProSerProSerProAlaProProSe 206  
 QY 121 TGTAGCGTGTGTCACGGTCTCTGGCACCTCTCCCTCGCAGTCCGAGC----- 169  
 Db 206 rProAlaProValProSerProAlaProSerProSerProSerProSerProAlaProPr 226  
 QY 170 -----TCTCAATGCGACAGGGGACCCAGGGCC----- 196  
 Db 226 oSerProSerProAlaProSerProSerProSerProSerProSerProSerProValPr 246  
 QY 197 ----TGGCCCTCCATCCAGAAAACCTAGACTCACCCTGCTGGAGCACCACCTCC 252  
 Db 246 oProSerProAlaProSerProAlaProSerProSerProSerProSerProAlaProPr 266  
 QY 253 GCAGCTTGGGAGCTATCTGAACTACTGGGCCCCCTTT---CAACGAGCCAGACT 309  
 Db 266 oProSerProProProProProProProProProProPheProAlaSerProMe 286  
 QY 310 TCACACCTCCCGCTGGGGGAGAGACTCTGCCAGGCCACTGTTGACTTGGAGGTGT 369  
 Db 286 tProSerProSerProProProProProProPro----- 297  
 QY 370 GCGGAGGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCTACAGCCACCTTC 429  
 Db 298 -AlaProThrProThrProThrProSerProSerProSerProSerProValProSe 317  
 QY 430 TGTGTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGTGAGCTGCGCGGAGCTGG 489  
 Db 317 rProAlaProValProProSerPro-----AlaProSerProAla 331  
 QY 490 CCCACTTGTGACAGCCTCCAGGCGCTGCTGGGAGCATTTGGGGGCTCATGGCAGCTC 549  
 Db 331 aPro-----SerProProProSerProAlaPro----- 340  
 QY 550 TGGGCTACCCACTGCCCCAGCGCTGCTGGGGAGTGAACCCACTTGGACTCTGGCCCTG 609  
 Db 341 -----ProThr-ProSerProSerProSerProSerProSerProSerProSerProS 358

```
QY 610 CCCACAGT 617
Db 358 erProSer 360

RESULT 15
DRPL RAT STANDARD; PRT; 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Atrophin-1 (Dentatorubral-pallidolysian atrophy protein).
GN DRPLA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.;
RT "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
RT homologue.";
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Epplen J.T., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
RT dentatorubral-pallidolysian atrophy (DRPLA) in rat.";
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -!- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC -!- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31777; AAA80337.1; -.
DR EMBL; X89453; CAA61623.1; -.
DR InterPro; IPR002951; Atrophin.
DR Pfam; PF03154; Atrophin-1; 2.
DR PRINTS; PR01222; ATROPHIN.
FT DOMAIN 165 171 POLY-PRO.
FT DOMAIN 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-PRO.
FT DOMAIN 477 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT DOMAIN 502 505 POLY-PRO.
FT DOMAIN 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT CONFLICT 455 455 N -> S (IN REF. 2).
FT CONFLICT 594 594 F -> L (IN REF. 2).
FT CONFLICT 689 689 P -> R (IN REF. 2).
FT CONFLICT 717 717 T -> M (IN REF. 2).
FT CONFLICT 737 737 A -> V (IN REF. 2).
FT CONFLICT 965 965 MISSING (IN REF. 2).
SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9B1F CRC64;
```

Alignment Scores:  
Pred. No.: 0.0341 Length: 1183

```
Score: 147.00 Matches: 78
Percent Similarity: 36.29% Conservative: 12
Best Local Similarity: 31.45% Mismatches: 66
Query Match: 10.08% Indels: 93
DB: 1 Gaps: 13

US-09-931-704-1 (1-797) x DRPL_RAT (1-1183)
QY 39 CCGCCAGCTCCGGAGAGGAGCGCCAGCCGCC----- 72
Db 502 ProPro-ProProGlyAlaTyrProHisProLeuHisProLeuHisProHisAlaHi 521
QY 73 -----GCCCCAGCCGCCAGCCGCC 89
Db 521 sProTyrAsnMetSerProSerLeuGlySerLeuArgProTyrProProGlyProAla-H 541
QY 90 ATGGACCTCCGAGCAGGAGGACTCTGGGGGATGTTAGCGTCTGTGACGGTGTCTCTGG 149
Db 541 isLeuProProSerHis-----GlyGlnValSerTyrSerGlnAlaGlyProAsnG 558
QY 150 CACCTCCCTGCGAGTCCAGCTCTCAATCGCA----- 180
Db 558 lyProProValSerSerSerSerSerSerSerGlySerSerSerGlnAlaAlaTyrSerC 578
QY 181 -----CAGGGAGCCCGAGGCGCTGGCC-----CCTCCATCCA 211
Db 578 ysSerHisProSerSerSerGlnGlyProGlnGlyAlaSerTyrProPheProValp 598
QY 212 GAAACCTATGACCTCACCCTACCTCGGAGCACC-----AACTCCG 253
Db 598 roProileThrThrSerSerAlaThrLeuSerThrValleAlaThrValAlaSerSerP 618
QY 254 CAGCTTGCTGGGACCTATCTGAACCTACTCGGGCCCCCTTTCAACGAGCCAGACTTCAA 313
Db 618 roAla-GlyTyrLysThrAlaSerProProGlyProPro-----GlnTyrSerLysArgAla 636
QY 314 CCTCC-----CCGCTGGGGGAGAGACTCTGCCAGGGC 349
Db 637 ProSerProGlySerTyrLysThrAlaThrProProGlyTyrLysProGly----- 653
QY 350 CACTGTTGACTGGAGGTGTGCGAGGCTCAATGACAACTCGCGCTACCCAGAACTA 409
Db 654 -----SerProProSerPheArgThrGlyThrProGlyTyr 666
QY 410 CGAGGCTACAGCCACCTTCTGTGTTACTTGGTGGCTCAACCGTCAGGCTGCCACTGC 469
Db 667 ArgGlyThrSerProPro-----AlaGlyPro-----GlyThrPheLys 679
QY 470 TGAGTGGCGCGCAG-----CCTGGCCACTTCTGACACCGCTCCAGGCGCTGCTGGG 523
Db 680 ProGlySerProThrValGlyProGlyProLeu-----ProProAlaGlyProSer--- 696
QY 524 CAGCATTGGGGCGTCATGGGAGCTCTGGGCTACCCACTGCCCGCCAGCCGCTGCTGGGAC 583
Db 697 -----SerLeuSerSerLeuProProProProAlaAla-ProThrTh 710
QY 584 TGAACCCACTTGGACT 599
Db 710 rGlyProProLeuThr 715
```

Search completed: January 27, 2003, 16:02:37  
Job time : 22.3445 secs

2

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:48:57 ; Search time 49.8794 Seconds  
(without alignments)  
6584.667 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 1458

Sequence: 1 attaaagctcgccgagcc.....tctctcttctgctccccccc 797

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09931704/runat\_27012003\_154126\_3635/app\_query.fasta.1.7189  
-DB=SPREMBL 21 -QWMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blotsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09931704 @CGN 1.1.633 @runat 27012003 154126 3635 -NCPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
1	1226	84.1	225	4	Q9UBD9			Q9ubd9 homo sapien

2	1193	81.8	225	11	Q9QZM3			Q9qzm3 mus musculus
3	177.5	12.2	1682	11	Q9QZR9			Q9qzr9 mus musculus
4	172.5	11.8	511	6	Q95JD0			Q95jd0 sus scrofa
5	172.5	11.8	566	6	Q95JD1			Q95jd1 sus scrofa
c	169.5	11.7	1953	5	Q9BIT7			Q9bit7 nephila mad
6	168.5	11.6	676	6	Q95JC9			Q95jc9 sus scrofa
7	166.5	11.4	889	16	Q9F2N5			Q9f2n5 streptomyce
c	166	11.5	444	5	Q9BIU6			Q9biu6 argiope tri
8	165	11.3	1532	4	Q02802			Q02802 homo sapien
9	164	11.3	399	5	Q9BIT8			Q9bit8 latrodectus
c	164	11.2	1497	4	Q9UMD9			Q9umd9 homo sapien
10	163	11.3	563	5	Q9BIT5			Q9bit5 nephila mad
11	161.5	11.1	326	10	Q02514			Q02514 santalum al
c	161.5	11.1	327	12	Q69145			Q69145 human herpe
12	161	11.0	1289	10	Q9FLO7			Q9flg7 arabidopsis
13	160	11.0	383	12	Q04357			Q04397 epstein-bar
14	159.5	10.9	1431	11	Q9JMH4			Q9jmh4 mesocricetu
15	159	11.0	373	5	Q9BIT9			Q9bit9 latrodectus
c	159	10.9	420	5	Q8SZ47			Q8sz47 drosophila
16	159	10.9	446	5	Q9VZC2			Q9vzc2 drosophila
17	158.5	11.0	2249	5	Q9NHM4			Q9nhm4 nephila cla
c	158	10.8	574	3	Q36027			Q36027 schizosacch
18	158	10.8	1838	11	Q88207			Q88207 mus musculus
19	157.5	10.8	1802	5	Q17163			Q17163 brugia mala
20	157	10.8	1497	4	Q9NQK9			Q9nqk9 homo sapien
21	157	10.8	1840	11	Q60467			Q60467 cricetus
22	156.5	10.7	325	5	Q17036			Q17036 caenorhabdi
c	156	10.8	636	5	Q16987			Q16987 araneus dia
23	156	10.7	1840	11	Q9JIO3			Q9ji03 rattus norv
24	155.5	10.7	319	5	Q17038			Q17038 caenorhabdi
25	155.5	10.8	871	5	O44358			O44358 nephila cla
c	155	10.6	437	12	Q69146			Q69146 human herpe
26	155	10.6	1835	13	Q91AU4			Q91au4 gallus gall
27	154.5	10.6	598	16	Q8VKN7			Q8vkn7 mycobacteri
28	152.5	10.5	890	5	Q77087			Q77087 alvinella p
29	151.5	10.5	410	5	Q16988			Q16988 araneus dia
c	151.5	10.4	1838	4	Q15094			Q15094 homo sapien
30	150.5	10.3	215	13	Q9PUJ2			Q9puj2 plethodon j
31	150.5	10.3	215	13	Q9PUJ1			Q9puj1 plethodon j
32	150.5	10.3	215	13	Q9PUJ0			Q9puj0 plethodon j
33	150.5	10.3	302	5	Q19079			Q19079 caenorhabdi
34	150.5	10.3	418	5	O01662			O01662 caenorhabdi
35	150	10.3	274	11	Q04154			Q04154 rattus norv
36	149.5	10.3	1433	11	Q07563			Q07563 mus musculus

## ALIGNMENTS

RESULT 1

Q9UBD9  
ID Q9UBD9 PRELIMINARY; PRT; 225 AA.  
AC Q9UBD9;  
DC 01-MAY-2000 (TREMELrel. 13, Created)  
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
DE Neutrophin-1/B-cell stimulating factor-3 (Cardiotrophin-like  
DE cytokine) (Similar to cardiotrophin-like cytokine,  
DE neutrophin-1/B-cell stimulating factor-3).  
GN CLC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]\_TaxID=9606;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99432254; PubMed=10500198;  
RA Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S.,  
RA Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,  
RA Simonet W.S., Boone T., Chang M.-S.,  
RT "Novel neutrophin-1/B cell-stimulating factor-3: A cytokine of the  
RT IL-6 family.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).  
RN [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=99382254; PubMed=10448081;  
RA Shi Y., Wang W., Youre P.A., Gohari S., Zukauskas D., Zhang J.,  
Ruben S., Alderson R.F.;  
RT "Computational EST database analysis identifies a novel member of the  
neuropoietic cytokine family";  
RL Biochem. Biophys. Res. Commun. 262:132-138 (1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Hu X., Xu Y., Zhang B., Peng X., Yuan J., Qiang B.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RA Strausberg R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR ENBL; AF176912; AAF00992.1; -;  
DR ENBL; AF172854; AAD54284.1; -;  
DR ENBL; AF176911; AAF00991.1; -;  
DR ENBL; AY049779; AAL15436.1; -;  
DR ENBL; BC012939; AAL12939.1; -;  
SQ SEQUENCE 225 AA; 25176 MW; E2DD4B6280833B55 CRC64;  
  
Alignment Scores:  
Pred. No.: 4.72e-87 Length: 225  
Score: 1226.00 Matches: 225  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 84.09% Indels: 0  
DB: 4 Gaps: 0  
  
US-09-931-704-1 (1-797) x Q9UBD9 (1-225)  
  
QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCGTGCTGTCGACGGTGCTCTGG 149  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
  
QY 150 CACTCCCTCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGGCCCTCCATC 209  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40  
  
QY 210 CAGAAACCTATGACCTACCGCTACCTGGAGCAGCACTCCGACGAGCTGGCTGGGACC 269  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
  
QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGACGACGACTTCAACCTCCCGCTGGGG 329  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
  
QY 330 GCAGAGACTCTGCCAGGCCACTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAAA 389  
Db 81 AlaGluThrLeuProArgAlaThrValAsnGluValTrpArgSerLeuAsnArg 100  
  
QY 390 CTGGGCTGAGGAGTGCAGCTGAGCTGGCGCGCAGCTTGTGTTACTTGGCGGCTC 449  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
  
QY 450 AACCGTCAAGTGCAGCTGAGCTGGCGCGCAGCTTGTGACGAGCTTGTGACGAGCTC 509  
Db 121 AsnArgGlnAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
  
QY 510 CAGGCTGTGGGAGCAGTTCGGGGCTCATGGCGGCTCATGGAGCTTGGGCTACCCAC 569  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
  
QY 570 CCGCTGCTGGGAGTGCAGCTGAGCTGGCGCGCAGCTTGTGACGAGCTTGTGACGAG 629  
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
  
QY 630 AAGATGACGAGCTTCTGGCTGCTGAAGGAGTGCAGAGCTGGGCTGGCGCTGGGCGCAAG 689  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 690 GACTTCACCGCTCAAGAGAGATGACGCTCCAGCAGCTGCGAGTCAACCTGACCTG 749  
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
  
QY 750 GGGGCTCATGGCTTC 764  
Db 221 GlyAlaHisGlyPhe 225  
  
RESULT 2  
Q9QZM3 PRELIMINARY; PRT; 225 AA.  
ID Q9QZM3;  
AC Q9QZM3;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Neurotrophin-1/B-cell stimulating factor-3.  
GN BSF3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99432254; PubMed=10500198;  
RA Senaldi G., Varnum B.C., Szamienito U., Starnes C., Lile J., Scully S.,  
RA Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,  
RA Simonet W.S., Boone T., Chang M.-S.;  
RT "Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the  
IL-6 family";  
RT Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463 (1999).  
RL EMBL; AF176913; AAF00993.1; -;  
DR MGD; MGI:193088; Bsf3.  
SQ SEQUENCE 225 AA; 25261 MW; 68B1FEAAB7F1A950 CRC64;  
  
Alignment Scores:  
Pred. No.: 1.72e-84 Length: 225  
Score: 1193.00 Matches: 218  
Percent Similarity: 98.22% Conservative: 3  
Best Local Similarity: 96.89% Mismatches: 4  
Query Match: 81.82% Indels: 0  
DB: 11 Gaps: 0  
  
US-09-931-704-1 (1-797) x Q9QZM3 (1-225)  
  
QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCGTGCTGTCGACGGTGCTCTGG 149  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
  
QY 150 CACTCCCTCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGGCCCTCCATC 209  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40  
  
QY 210 CAGAAACCTATGACCTACCGCTACCTGGAGCAGCACTCCGACGAGCTGGCTGGGACC 269  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
  
QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGACGACGACTTCAACCTCCCGCTGGGG 329  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
  
QY 330 GCAGAGACTCTGCCAGGCCACTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAAA 389  
Db 81 AlaGluThrLeuProArgAlaThrValAsnGluValTrpArgSerLeuAsnArg 100  
  
QY 390 CTGGGCTGAGGAGTGCAGCTGAGCTGGCGCGCAGCTTGTGTTACTTGGCGGCTC 449  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
  
QY 450 AACCGTCAAGTGCAGCTGAGCTGGCGCGCAGCTTGTGACGAGCTTGTGACGAGCTC 509  
Db 121 AsnArgGlnAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
  
QY 510 CAGGCTGTGGGAGCAGTTCGGGGCTCATGGCGGCTCATGGAGCTTGGGCTACCCAC 569  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
  
QY 570 CCGCTGCTGGGAGTGCAGCTGAGCTGGCGCGCAGCTTGTGACGAGCTTGTGACGAG 629  
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
  
QY 630 AAGATGACGAGCTTCTGGCTGCTGAAGGAGTGCAGAGCTGGGCTGGCGCTGGGCGCAAG 689  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200



```

Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
QY 570 CGCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCACAGTACTTCTCCAG 629
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AAGATGGAGGACTTCTGGCTGCTGAAGAGCTGCAGACCTGGCTGGGCTCGGCCAAG 689
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 690 GACTTCAACCGCTCAAGAGAGATGCAGCTCCAGCTGCAGTGCAGTCCACTGCACCTG 749
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
QY 750 GGGGCTCATGGCTTC 764
Db 221 GluAlaHisGlyPhe 225

RESULT 3
Q9QZR9 PRELIMINARY; PRT; 1682 AA.
AC Q9QZR9;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Alpha 4 collagen IV.
GN COL4A4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
EX MEDLINE=20005934; PubMed=10534397;
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
RA Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;
RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
RT mouse model of alport syndrome.";
RL Genomics 61:113-124(1999).
DR EMBL; AF169388; AAD50450.1; -.
DR MGD; MGI:104687; Col4a4.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001442; ProcollagenC4.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 22.
DR ProDom; PD000007; Collagen; 4.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1682 AA; 164096 MW; 6F7B679EDD76E904 CRC64;

Alignment Scores:
Pred. No.: 1.66e-05 Length: 1682
Score: 177.50 Matches: 77
Percent Similarity: 37.60% Conservative: 14
Best Local Similarity: 31.82% Mismatches: 81
Query Match: 12.17% Indels: 70
DB: 11 Gaps: 13

US-09-931-704-1 (1-797) x Q9QZR9 (1-1682)
QY 12 GCGGAGCGCGGCTCGCCCT-----CCCACTCGGCAGCTCCGGGAGA 56
Db 1235 AlaGlyAlaProGlyArgAlaAlaLysGlyAspIleProAspPro-GlyProProGlyAs 1254
QY 57 GGAG-----CCGACCCCGCGCGCGCCAGCCCGCCAGCCCGCCATGACCTCCGAGC----- 103
Db 1254 pArgGlyProProGlyProAspGlyProArgGlyValPro-GlyProProGlySerProG 1274
QY 104 -----AGGGAGCTGTGGGGATGTAGTGCCT----- 133

```

```

Db 1274 LysnValAspLeuLeuLysGlyAspProGlyAspCysGlyLeuProGlyProProGly 1294
QY 134 -----GTGCACGGTGTCTTGGCACCTTCCCTGTCAGTGCAGCTCTCAAT 176
Db 1294 eArgGlyProProGlyProGlyCysGlnGlyProGlyProGlyCysAspGlyLysAspG 1314
QY 177 GCGACAGGGAGCCAGGGCCCTGCCCTCCATCCAGAAACCACTATGACCTACCCGCTAC 236
Db 1314 LysGlnLysGlyProMetGlyLeuProGlyLeuProGly-----ProGlyLeuP 1331
QY 237 CTGGAGCACCACCTCCGAGCTTGGCTGGGACCTATCTCAACTACTGGGCCCCCTTTC 296
Db 1331 roGlyAlaProGlyLys-----GlyLeuProGlyProProGly 1345
QY 297 AACGA-----GCCAGACTTCAACCTCCCGCTCCGCGGCGAGAG 335
Db 1345 rGlyGlyProValGlyProProGlyCysArgGlyGluProGlyProProAlaAspVal 1365
QY 336 ACTTGTCCCGAGGCGCACTTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAACTGCGG 395
Db 1365 spSerCysProArgIleProGlyLeuProGlyValProGlyProArgGlyProGluGly 1385
QY 396 CT-----GACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACTTGGCTGCGCTC 449
Db 1385 laMetGlyGluProGlyArgArgGlyLeuProGlyPro----- 1397
QY 450 AACCGTCCAGGCTGCCACTGCTGAGCTGCGCGCGAGCTGCGCCA----- 493
Db 1398 -----GlyCys-----LysGlyGluProGlyProAspGlyArgArgGlyG 1411
QY 494 -----CTTGTGACCAAGCTCCAGGCGCTGCTGGGCGAGCATTCGCGCGCTCATGGCAGCT 548
Db 1411 lnAspGlyIleProGlySerProGlyProGlyProGlyArgLys--GlyAspThrGlyGlu 1430
QY 549 CTGGGCTACCACTGCCCGCGCGCTGCTGGAGTGAACCCACTTGGACTCTGGGCTT 608
Db 1430 laGlyCysProGlyAlaProGlyPro-ProGly-----ProThrGlyAspProGlyPro 1447

RESULT 4
Q95JD0 PRELIMINARY; PRT; 511 AA.
AC Q95JD0;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Basic proline-rich protein.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]_TaxID=9823;
RP SEQUENCE FROM N.A.
RC TISSUE=PAROTID GLAND;
RA Zhang Q., Szalay A.A., Kyeyune-Nyombi E., Sands J.F., Oberg K.C.,
RA Tieche J.-M., Leonora J.;
RT "Cloning and expression of a novel proline-rich protein from porcine
RT parotid glands.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY035848; AAK61382.1; -.
DR PRINTS; PRO1582; KV33CHANNEL.
SQ SEQUENCE 511 AA; 48483 MW; AB04597964C448D7 CRC64;

Alignment Scores:
Pred. No.: 3.27e-05 Length: 511
Score: 172.50 Matches: 72
Percent Similarity: 37.31% Conservative: 3
Best Local Similarity: 35.82% Mismatches: 79
Query Match: 11.83% Indels: 47
DB: 6 Gaps: 9

US-09-931-704-1 (1-797) x Q95JD0 (1-511)
QY 30 CTTCCCACTCCGCGAGCTTCCGGGAGGAGCGGCAC-----CCGCGCGCGCCAGC 80

```

```

Db 285 ProProGlyProPro- ProProGlyProAlaProHisGlyAlaArgPro- ProProGlyP 304
QY 81 CCAGAGCCCATGGACCTCCAGAGGAGGACTCGTGGGGATGTTAGCGTGTGCAGC 140
Db 304 roProProGlyProPro- ProProGly- 317
QY 141 GTGCTCTGGACCTCCCTGAGTCCAGCTCTCAATCGCACAGGAGGAGCCAGGCGCTGC 200
Db 317 roGlyAlaArgProProGly- ProProProGlyProPro- ProProGlyPro- Al 336
QY 201 CCTCCATCCAGAAAACCTATGACTACCCGCTACCTGGAGCACCACCACTCCGAGCTTG 260
Db 336 aProProGlyAlaArgProPro- ProGlyProProProGlyProPro- 353
QY 261 GCTGGGAGCTATCTGAACCTACCTGGGCGCCCTTTCAACGAGCCAGACTTCAACCCCTCC 320
Db 354 -----ProGlyProAlaProProGlyAlaArgProProGlyPro- 369
QY 321 CGCTGGGGCAGAGACTCTGCCAGGCGCACTGTTGACTTGGAGGTGTGGCGAAGCCTC 380
Db 369 roProGlyProProPro- 381
QY 381 AATGACAACTGGCGCTGACCCAGAACTAGAGGCTCAGACCACCTTCTGTACTTG 440
Db 381 lyAlaArgProProGlyProPro- ProProGlyProPro- 397
QY 441 CGTGGCTCAACCTGAGGCTGCCACTGCTGAGCTGCGCGCAGCTGCCACCTTCTGC 500
Db 398 -----ProAlaProProGly- 409
QY 501 ACCAGCTCCAGGCGCTGCTGGGAGCATTGCGGGCGTCATGGCAGCTCTGGGCTACCCA 560
Db 410 --ProProGlyProPro- ProProGlyProAlaProProGlyAlaArgProLeuPro- 428
QY 561 CTGCCCCAGCGCTGCTGGAGTGAACCCACTTGGACTCTGGCCCTGCC 611
Db 429 --GlyPro- ProProGlyProPro- 441

RESULT 5
Q95JD1 PRELIMINARY; PRT; 566 AA.
AC Q95JD1; (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Basic proline-rich protein.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PAROTID GLAND;
RA Zhang Q., Szalay A.A., Kyeyune-Nyombi E., Sands J.F., Oberg K.C.,
RA Tieche J.-M., Leonora J.;
RT "Cloning and expression of a novel proline-rich protein from porcine
RT parotid glands."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY035847; AAK61381.1; -.
DR PRINTS; PR01582; KV33CHANNEL.
SQ SEQUENCE 566 AA; 53213 MW; E33B3B5E1BDEB81A CRC64;

Alignment Scores:
Pred. No.: 3.34e-05 Length: 566
Score: 172.50 Matches: 72
Percent Similarity: 37.31% Conservative: 3
Best Local Similarity: 35.82% Mismatches: 79
Query Match: 11.83% Indels: 47
DB: 6 Gaps: 9

US-09-931-704-1 (1-797) x Q95JD1 (1-566)

```

```

QY 30 CTCTCCACTCCCGAGCTCTCGGAGAGAGCGGCAC-----CCGCGCGCCCGCAGC 80
Db 285 ProProGlyProPro- ProProGlyProAlaProHisGlyAlaArgPro- ProProGlyP 304
QY 81 CCCAGCCCCATGGACCTCCAGAGGAGGACTCGTGGGGATGTTAGCGTGTGCAGC 140
Db 304 roProProGlyProPro- ProProGly- 317
QY 141 GTGCTCTGGACCTCCCTGAGTCCAGCTCTCAATCGCACAGGAGGAGCCAGGCGCTGC 200
Db 317 roGlyAlaArgProProGly- ProProProGlyProPro- ProProGlyPro- Al 336
QY 201 CCTCCATCCAGAAAACCTATGACTACCCGCTACCTGGAGCACCACCACTCCGAGCTTG 260
Db 336 aProProGlyAlaArgProPro- ProGlyProProProGlyProPro- 353
QY 261 GCTGGGAGCTATCTGAACCTACCTGGGCGCCCTTTCAACGAGCCAGACTTCAACCCCTCC 320
Db 354 -----ProGlyProAlaProProGlyAlaArgProProGlyPro- 369
QY 321 CGCTGGGGCAGAGACTCTGCCAGGCGCACTGTTGACTTGGAGGTGTGGCGAAGCCTC 380
Db 369 roProGlyProProPro- 381
QY 381 AATGACAACTGGCGCTGACCCAGAACTAGAGGCTCAGACCACCTTCTGTACTTG 440
Db 381 lyAlaArgProProGlyProPro- ProProGlyProPro- 397
QY 441 CGTGGCTCAACCTGAGGCTGCCACTGCTGAGCTGCGCGCAGCTGCCACCTTCTGC 500
Db 398 -----ProAlaProProGly- 409
QY 501 ACCAGCTCCAGGCGCTGCTGGGAGCATTGCGGGCGTCATGGCAGCTCTGGGCTACCCA 560
Db 410 --ProProGlyProPro- ProProGlyProAlaProProGlyAlaArgProLeuPro- 428
QY 561 CTGCCCCAGCGCTGCTGGAGTGAACCCACTTGGACTCTGGCCCTGCC 611
Db 429 --GlyPro- ProProGlyProPro- 441

RESULT 6
Q9BIT7 PRELIMINARY; PRT; 1953 AA.
AC Q9BIT7;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Major ampullate spideroin 2-like protein (Fragment).
OS Nephila madagascariensis.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=115969;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21173804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
RT Fibroin Sequences."
RL Science 291:2603-2605 (2001).
DR EMBL; AF350276; AAK30605.1; -.
DR NON_TER 1
FT NON_TER 1953
SQ SEQUENCE 1953 AA; 159383 MW; C82B4DACD043C3BD CRC64;

Alignment Scores:
Pred. No.: 7.14e-05 Length: 1953
Score: 169.50 Matches: 87
Percent Similarity: 34.88% Conservative: 18
Best Local Similarity: 28.90% Mismatches: 106
Query Match: 11.72% Indels: 90
DB: 5 Gaps: 15

US-09-931-704-1 (1-797) x Q9BIT7 (1-1953)

```





```
Query Match: 11.48% Indels: 77
DB: 5 Gaps: 16
US-09-931-704-1 (1-797) x Q9BIU6 (1-444)
QY 750 CCAGTGCAGGTGACTGACAGCTGTGAGGCTGCATCTTCTTGTAGCCGGTTGAAGT 691
Db 19 ProGlySerGly-----GlyGlnGlnGlyProGlyGlyGlnGly 31
QY 690 CCTTGG---CCGAGCGCCACAGCAGGTCTGCAGCT-----CCTTCA 652
Db 32 ProGlyGlyProSerAlaAlaAlaAlaAlaAlaAlaGlyProGlyTyrGlyProGly 51
QY 651 GCAGCCAGCAAGTCTCTTCTTCTGAGGAAGTCACTGTGGCAGCGCCAGGAGTCCCAAG 592
Db 52 AlaGlyGlnGlnGlyProGlySerGlyGlyGlnGlnGlyGlyGlnGlnGln 71
QY 591 TGG----- 589
Db 72 GlyProGlyGlyAlaGlyGlnGlyGlyProArgGlyGlnGlyProTyrGlyProGlyAla 91
QY 588 GTTCAGTCCAGCGAGCGCTGGGCGAGTGGGTAGCCAGAGCTGCCATGACGCCGCA 529
Db 92 AlaAlaAlaAlaAlaAlaAlaGlyTyrGly-----ProGly 104
QY 528 TGCTGCCAGCAGCCCTGGAGGCTGGTGACAGAGTGGCCAGGCTGGCGGCAGCTCAG 469
Db 105 AlaGlyGlnGlnGlyProGlySer-----GlnGlyProGlySerGlyGlyGlnGln 121
QY 468 CAGTGGCAGCTGACGGTTGAGGCCACGCAAGTAAACAGAGGTGGCTGTAGGCTGT 409
Db 122 -----GlyProGlySerGlnGlyProTyrGlyProSer 132
QY 408 AGTTCTGGTCCAGCGCAGTGTTCATTGAGGCTTCGCCACACCTCCCAAGTCAACAGTGG 349
Db 133 AlaAlaAlaAlaAlaAlaAlaGlyProGlyTyrGlyProGlyAlaGlyGlnGlnGly 152
QY 348 CCTTGGCAGAG-----TCTCTGCCCCAGCGGGGAGGGT----- 313
Db 153 ProGlySerGlnGlyProGlySerGlyGlyGlnGlnGlyProGlyGlyGlnGlyProTyr 172
QY 312 -----TGAAGTCTGGCTGTTGAAGGGGGGCCAGGTAGTTCAGATAGTCCCG 262
Db 173 GlyProSerAspAlaAlaAlaAlaAlaAlaGlyProGlyTyrGly----- 188
QY 261 CCAAGCTCGGAGTGGTGCTCCAGGTAGCGGTGAGGTCTAGGTCTTCTGATGGAGG 202
Db 189 ProGlyAlaGlyGlnGlnGlyProGlySerGly-----GlyGlnGlnGlyGlyGlnGlySer 207
QY 201 GGCCAGGCGCTGGTCCCTGTGCGATTGAGAGCTGGCAGTGGCAGGGAGGTGCCAGAGCA 142
Db 208 GlyGlnGln---GlyPro-----GlyGlyAlaGlyGlnGlyGlyProArgGly 222
QY 141 CGGTGCACA---GGCAGCGTAACTATCCCCAGTCCCTGCTGGAGGTCCATGGGGC 85
Db 223 GlnGlyProTyrGlyProGlyAlaAlaAlaAlaAlaAlaAlaGlyTyrGlyPr 242
QY 84 TGGGCTGGCGCGCGGTGGGCTCC---TCTCCCGA---GGCTGGCGGAGTGGGAG 31
Db 242 oGlyAlaGlyGlnGlnGlyProGlySerGlnGlnGlyProGlySerGlyGlyGlnGlnGlyPr 262
QY 30 GCGCAGCGCGCGC 18
Db 262 oGlySerGlnGly 266
RESULT 10
Q02802 PRELIMINARY; PRT; 1532 AA.
AC Q02802;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Autoantigen (Fragment).
```

GN BP180.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=FORESKIN;  
RX MEDLINE=92381323; PubMed=1324962;  
RA Giudice G.J., Emery D.J., Diaz L.A.;  
RT "Cloning and primary structural analysis of the Bullous pemphigoid  
autoantigen, BP180.";  
RL J. Invest. Dermatol. 99:243-250(1992).  
DR EMBL; M91669; AAA35605.1; -.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF01391; Collagen; 8.  
FT NON TER 1  
SQ SEQUENCE 1532 AA; 154568 MW; BF35054CF93BBEE3 CRC64;

Alignment Scores:  
Pred. No.: 0.000153 Length: 1532  
Score: 165.00 Matches: 77  
Percent Similarity: 29.41% Conservative: 23  
Best Local Similarity: 22.65% Mismatches: 96  
Query Match: 11.32% Indels: 144  
DB: 4 Gaps: 13

US-09-931-704-1 (1-797) x Q02802 (1-1532)

QY 20 CCGCGCTCGCCCTCCCACTCCGCCAGCTCCGGGAGAGAGCGCCACCCGGCGGCCAG 79  
Db 667 ArgGlyGluAlaGlyProProGlySerGlyGlyGluValGlyGlyAlaAlaGlyGlu 686  
QY 80 CCCAGCCCCATGGACCTCCGAGC----- 103  
Db 687 ProGlyProHisGlyProProGlyValProGlySerValGlyProGlySerSerGly 706  
QY 104 -----AGGGGACTCGTGGGGATGTT 124  
Db 707 SerProGlyProGlnGlyProProGlyProValGlyLeuGlnGlyLeuArgGlyGluVal 726  
QY 125 AGCTGCGCT-----GTGCACGGTGTCTGTGGACCTCCCTGCGAGTGCACCTCT 172  
Db 727 GlyLeuProGlyValGlySerGlyAspGlyProMetGlyProProGlyProGlyProGlyAsp 746  
QY 173 CAATCGCACAGGGGACCCAGGSCCTGGCCCTCCATCCAGAAAACCTATGACCTCACCCG 232  
Db 747 GlnGlyGluGlySerGlyProArgGlyLeuThrGlyGluProGly-----MetArgGly 763  
QY 233 CTACCTGAGGACCACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTCTGGGCCCCC 292  
Db 764 LeuProGlyAla-----ValGlyGluPro 771  
QY 293 TTTCAACAGCCAGACTTCAACCTCCCGCTGGGGGAGAGAGACTCTCCAGGGGCCAC 352  
Db 772 GlyAlaGlySerGlyAlaMetGlyProAlaGlyProAspGlyHisGlnGlyProArgGlyGlu 791  
QY 353 GTTGTGACTGGAGGTGGCGAAGCTCAATGACAACTGCGGCTGACCCAGAACTACCA 412  
Db 792 GlnGlyLeuThrGlyWet-----ProGlyIleArg 801  
QY 413 GGCCTACAGCCACCTTCTGTGTACTTGTGGCTCCCAACCGCTCAGGCTGCCACTGCTGA 472  
Db 802 Gly-----ProProGly-----ProSerGly----- 808  
QY 473 GCTGCGCGCAGCTGGC---CCACTTGTGACACGACCTCCAGGGCTCTCTGGG----- 523  
Db 809 -----AspProGlyLysProGlyLeuThrGlyProGlnGlyProGlnGlyLeuPro 825  
QY 524 CAGCATTGGCGGCTCATGGCAGCTCTGGCTTACCACTGCCCCAGCCGCTGCCTGGGAC 583  
Db 826 GlyThrProGlyArgProGlyIleGlyGlyGluProGlyAlaPro----- 840





Best Local Similarity: 30.96% Mismatches: 78  
Query Match: 11.27% Indels: 71  
DB: 5 Gaps: 14

US-09-931-704-1 (1-797) x Q9B1T5 (1-563)

```
QY 684 CCGAGCCGACGAGCGTCTGACAGTCTCTTCCAGCAGCAGAGTCTCTTCTGGA 625
DB 255 ProGlySerAlaAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyProGly 274
QY 624 GGAAGTCACTGTGGCAG-----GGCAGAGTCCAAAGTGGTTCAGTCCAGCACA 574
DB 275 GlyTyrGlyProGlyGlnGlnGlyProGlyGlnGlnGlyProSerGlyProGlySerAla 294
QY 573 GCGGCTGGGAGTGGTAGCCAGAGCTGCCATGACGCCCGCATGTGTGCCAGCAGGC 514
DB 295 AlaAlaAlaAlaAlaAlaGlyProGly-----ProGlnGly 306
QY 513 CTGGAGGCTGGTGCAGAGTGGCCAGGCTGCGCGCAGCTCAGCAGTGGCAGCCTGAC 454
DB 307 ProGlyGlyTyr-----GlyProGly-----GlnGlnGlyPro--- 317
QY 453 GGTGAGGCCACGCAAGTAACACAGAGTGGCTGTAGCTGTGAGTTCGTGGGTACGC 394
DB 318 -----GlyGlyTyrGlyProSerGlyProGlySerAla 328
QY 393 GCAGTTTGTCTAGGCTTCGCCACACCTCCAAAGTCAACAGTGGCCCTGGGCAGAGTCT 334
DB 329 AlaAlaAla-----AlaAlaAlaGlyProGlyGlnGln----- 340
QY 333 CTGCCCCAGCGGGAGGTTGAAGTCTGGCTGTGAAAGGGG-----GGCCCCAGGTAGT 277
DB 341 --GlyProGlyGlyTyrGlyProGlyGlnGlnArgProSerGlyTyrProGlyGln 359
QY 276 TCAGATAGTCCACAGCAAGCTGGGAGTGTGCTCCAGGTAGCGGTGAGGTCAATAGG 217
DB 360 GlnGlyProSerGlyProGlySerAlaAlaAlaAlaAlaGlyProGlyGlnGln 379
QY 216 TTTTCTGATGGAGGGGCGAGCGCTGGGTGCCCTGTGCGATTGAGAGCTGCGACTGCAG 157
DB 380 GlyProGlyAlaTyrGlyProSer---GlyPro----- 389
QY 156 GGAGGTGCCAGAGCACCGTGCACAGGACGCTAATATCCCCACAGATCCCTGTCTCGGA 97
DB 390 --GlySerAlaAlaAlaAlaAlaGly-----LeuGly 399
QY 96 GGTCCATGGG---CTGGGGTGGCGCGCGCGGTGGCGCTCTCT----- 54
DB 400 Gly-TyrGlyProAlaGlnGlnGlyProSerGlyAlaGlySerAlaAlaAlaAlaAla 419
QY 53 -----CCCGAGGCTGG-----CGGAGTGGGAGGGGAGCGCGGCTCC 15
DB 419 aAlaGlyProGlyGlyTyrGlyProValGlnGlnGlyProSerGlyProGlySer 437
```

RESULT 14

```
O22514
ID O22514 PRELIMINARY; PRT; 326 AA.
AC O22514;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE Proline rich protein.
OS Santalum album (white sandalwood).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Santalales; Santalaceae; Santalum.
OX NCBI_TaxID=35974;
RN [1]
RP SEQUENCE FROM N.A.
RA Bhattacharya A., Sita L.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF020261; AAB70928.1;
SQ SEQUENCE 326 AA; 33407 MW; B10011C45A86DCD7 CRC64;
```

Alignment Scores: 0.000216 Length: 326  
Pred. No.: 161.50 Matches: 63  
Score: 36.87% Conservative: 17  
Best Local Similarity: 29.03% Mismatches: 70  
Query Match: 11.08% Indels: 67  
DB: 10 Gaps: 11

US-09-931-704-1 (1-797) x O22514 (1-326)

```
QY 7 GCTTCGCGGAGCCGCGCTCGCTCCACTCCGACGCTCCGCGAGCTCCG-----GGAGAGGAG 60
DB 152 AlaThrProThrProArgProSerSerProThrSerThrProThrAspGlyGluThr 171
QY 61 CCGCACCAGCGCGCG-----CCAGCCCGAGCCCGACCTCCGAGCAGGGGACT 111
DB 172 ProArgProProGlyProArgSerProSerPro-GlyProProSer----- 188
QY 112 CGTGGGGATGTTAGCGTGCCTGTGCACGGTCTCTGGCACCTCCCTGCACTGTCAGTGCAGCTC 171
DB 189 -----CysSerProSerProLysSerProSerProProAlaSerSerProPr 204
QY 172 TCATCGCACAGGAGCCAGCGCTGCGCCCTCCATCCAGAAAACCTATGACCTCACC 231
DB 204 oArgSerArgProGlyProProAspTyrThrThrSerProSerProProThrProArgSe 224
QY 232 GCTACTGGAGCACCACCACTCCGACGCTTGGCTGGGACCTATCTGAACCTACTGCGGCC 291
DB 224 rValPro-----ProThrProProAlaSer-----ProSerProPr 236
QY 292 CTTTCAACGAGCAGACTTCAACCTCCCGCTCCCGGGGAGAGACTCTGCCAGGGCCA 351
DB 236 O-----ThrAlaLysProSerProProSerArgGlySerSer----- 248
QY 352 CTGTTGACTTGGAGTGTGGGAAAGCTCAATGACAACTGCGGCTGACCCAGAACTACG 411
DB 249 -----ProSerProProThrSerProProThrProPr 260
QY 412 AGGCTACAGCCACTTCTGTG----- 433
DB 260 g-----ProProSerTyrSerProSerProThrProProSerSerArgProSerPr 277
QY 434 -----TTACTTTCGCTGGGCTCAACCGCTCAGCGCTGCGACCTGCTGAGCTGC 477
DB 277 oProLeuArgSerProIleLeuThrProSerProAla-----AlaVa 292
QY 478 GCGCAGCCTGCGCCACTTCTGCACCGCTCCAGCGCTGCTGGGAGCAGCATTCGCGGCG 537
DB 292 lPro-----ProIleGlyArgSerProProSerProIleAspProProCysSerSe 309
QY 538 TCATGGCAGCTTGGGCTACCCACTGCCCGCGCTGCTGGGACT 584
DB 309 rProGluProSerProProThrSerPro-ProThrProSerSer 324
```

RESULT 15

```
O69145
ID O69145 PRELIMINARY; PRT; 327 AA.
AC O69145;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE The first atg start codon is the AA before the stop codon in ORF1
DE (Fragment).
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=P3HR-1;
RX MEDLINE=87198876; PubMed=3033277;
RA Jensen H.B., Farrell P.J., Miller G.;
```



RT "Sequences of the Epstein-Barr Virus (EBV) large internal repeat form  
RT the center of a 16-kilobase-pair palindrome of EBV (P3HR-1)  
RT heterogeneous DNA [published erratum appears in J Virol 1987  
RT Sep;61(9):2950]."; 1506(1987).  
RL J. Virol. 61:1495-1506(1987).  
DR EMBL; M15972; AAA66541.1; -.  
FT NON TER 1  
SQ SEQUENCE 327 AA; 33982 MW; 84EAF9F5D4ED97A CRC64;

## Alignment Scores:

Pred. No.:	0.000216	Length:	327
Score:	161.50	Matches:	81
Percent Similarity:	37.07%	Conservative:	15
Best Local Similarity:	31.27%	Mismatches:	72
Query Match:	11.08%	Indels:	91
DB:	12	Gaps:	18

US-09-931-704-1 (1-797) x Q69145 (1-327)

QY	13	CCGGAGCGCGCTCCCTCCACTCCGCCAGCCT-----	48
DB	10	ProAspProGlySerAlaSerProAlaAspArgProHisSerGlyArgLeuLeuGlyAla	29
QY	49	-----CCGGAGAGGAGCGCGCACCCGGCGGC	75
DB	30	SerArgArgGlyTyrPheCysProSerLeuCysProSerGluGluProGlyThrSerGly	49
QY	76	CCAGCCCGAGCCCATGACCTCCGAGCAGG-----GGACTCGTGGGGGATGTTAG	126
DB	50	-ThrProGluProLeuGlyProAlaSerArgProGlyLeuArgSerProLeuSe	69
QY	127	CCTGCTCTGCAC-----GGTGTCTGGCACCTCC	156
DB	69	r----ProVallysProLysGluCysLeuArgGlyAlaThrLeuGlyAlaGlnAlaPro--	87
QY	157	CTGCAGTCCAGCTCTCAATGCAC-----AG	183
DB	88	---GluSerArgGlyGlnGlyHisLeuArgValProProArgValProGlyGlnProG1	106
QY	184	GGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATACCTACCGCTACCTGGAGC	243
DB	106	uGlyProArgGln---ProGlyArgProGlnArgProValProArgProPheProGlyue	125
QY	244	ACCAACTCCG-----CAGCTTGGCTGGGACCTATCTGAACCTACCTGG	285
DB	125	uGlnSerProGlyCysProProGluGlyThrLeuGly-----ValProSe	140
QY	286	GGCCCTTTTCAACGAGCGAGACTTCAACCTCCCGCTGGGGCGAGAGACTCTGCCCA	345
DB	140	rProProLeuGln---AlaArgAlaSerProSerArgArgGlyAlaSer-----	155
QY	346	GGGCCACTCTTACCTTGGAGGTGTGGCGAAGCTCAATGACAACTGGCGCTGACCCAGA	405
DB	156	-----LeuGly-----ProHisValGlnProHisArgAspProSe	167
QY	406	ACTACGAGCCCTACAGCCACTTCTGTGTACTTGGTGGCCTCAACCGCTCAGGCTGCCA	465
DB	167	r-----GlyProAspProThr-----GlyProSerLeuCysPr	179
QY	466	CTGCTGAGCTGGCGCG-----CAGCTGGCCCACTTCTGCACAGCCTCC	510
DB	179	o-----ProAlaProLeuGlnProSerLeuHisProArgProGlnLeuLeuAlaSerPr	197
QY	511	AGGGCTGTCTGGGAGCATTTGGGG-----CGTCATGGCAGCTCTGGGTACCCACTGC	564
DB	197	oGlyProProGlyGlnProGluGlyProArgGlnProGly-ArgValAlaPheProLeuP	217
QY	565	CCAGCG---CTGCTGGGAGTGAACCCACTTGGACTCTGCGCCCTGCCAC	614
DB	217	roTriProLeuLeuProAlaSerHisProSerProLeuSerLeuProProHis	234

Search completed: January 27, 2003, 16:16:41

Job time : 59.8794 secs





KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder.

XX Homo sapiens.

OS WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

PR 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSRQ INC.

XX Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-457603/49.

DR N-PSDB; AAH99772.

XX Isolated human polynucleotides encoding polypeptides, useful for the  
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 PT

XX Claim 20; Page 278; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 CC AAM25963. The proteins can have activities based on the tissues and  
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;  
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,  
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders.

XX SQ Sequence 253 AA;

Alignment Scores:

Pred. No.: 2,41e-104 Length: 253  
 Score: 1364.00 Matches: 251  
 Percent Similarity: 99.21% Conservative: 0  
 Best Local Similarity: 99.21% Mismatches: 2  
 Query Match: 93.55% Indels: 0  
 DB: 22 Gaps: 0

US-09-931-704-1 (1-797) x AAM25831 (1-253)

QY 6 AGCTTCGGAGCGGGGCTCGCCCTCCCACTCGCGAGAGGAGGAGCGCA 65

Db 1 SerPheAlaGlyAlaAlaArgProSerThrProProAlaSerGlyArgGlyAlaAla 20

QY 66 CCGCGCGCGCCAGCCCGCATGACCTCGAGCAGGGGACTCGTGGGGATGTTA 125

Db 21 ProGlyArgProGlyProSerProMetAspPueuArgAlaGlyAspSerTrpGlyMetLeu 40

QY 126 CGGTCCCTGTGACCGGTGCTGTGGCACCTCTCCCTGCGAGTCCGAGCTCTCAATCGCAGGG 185

Db 41 AlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAsnArgThrGly 60

QY 186 GACCAGAGGCGGCGCTCCATCCAGAAACCTATGACCTACCCGCTACCTGGAGCAC 245

Db 61 AspProGlyProGlyProSerIleGlnIysThrTyrAspLeuThrArgTyrLeuGluHis 80

QY 246 CAATCCCGCAGCTTGGCTGGGACCTATCTGAACTACTCTGGGCCCCCTTTTCAACGAGCCA 305

Db 81 GlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPheAsnGluPro 100

QY 306 GACTTCAACCTCCCGCTGGGCGAGAGACTCTGCCAGAGGCGCACTGTGTGCTGGAG 365

Db 101 AspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeuGlu 120

QY 366 CTGTGGCGAAGCCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCTACAGCCAC 425

Db 121 ValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSerHis 140

QY 426 TTTCTGTGTACTTGTGGCTCTCAACCTGAGGCTGCCACTGCTGAGTGCCTCGCGCAGC 485

Db 141 LeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgSer 160

QY 486 CTGGGCGGCTTGTGACCAAGCTCCAGAGGCTGCTGGGCGAGCATTTGGGCGCTCATGGCA 545

Db 161 LeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMetAla 180

QY 546 GCTCTGGGCTACCACTGCGGCGGCTGCTGGGAGCTGACCCACTTGGACTTGGCTCTGGC 605

Db 181 AlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrProGly 200

QY 606 CTGCGCCACAGTACTTCTCCAGAAAGATGGAGCACTTCTGGCTGTGGAAGAGCTCGAG 665

Db 201 ProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLysGluLeuGln 220

QY 666 ACCTCGCTGTGGCGCTCGGCCAAGACTTCAACCCGCTCAAGAGAGAGATGCGAGCTTCCA 725

Db 221 ThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnProPro 240

QY 726 GCAGCTGCGAGTCACTGCGGCTGCGGCTCATGGCTTC 764

Db 241 AlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 253

RESULT 2

AAW29715

ID AAW29715 standard; Protein; 225 AA.

XX AAW29715;

AC AAW29715;

DT 09-NOV-1998 (first entry)

XX Human neurotrophic factor NNT-1.

XX NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;

KW Alzheimer's disease; Parkinson's disease; Huntington's disease;

KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;

KW peripheral neuropathy; dystrophy; neural retina degeneration;

KW common variable immunodeficiency; CVID; selective IgA deficiency;

KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;

XX Homo sapiens.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..27

FT Peptide /label= Sig\_peptide

FT Protein 28..225

FT Protein /label= Mat\_protein

XX WO9833922-A1.

PN WO9833922-A1.

XX 06-AUG-1998.

PD 06-AUG-1998.

XX 02-FEB-1998; 98WO-US02363.  
 PF XX  
 PR 30-JAN-1998; 98US-0016534.  
 PR 03-FEB-1997; 97US-0792019.  
 XX  
 XX (AMGE-) AMGEN INC.  
 PA  
 PA Chang M, Elliot GS, Sarmiento U, Senaldi G;  
 PI  
 PI WPI; 1998-437475/37.  
 DR N-PSDB; AAV47510-11.  
 DR  
 XX Newly isolated nucleic acid encoding human or murine neurotrophic  
 PT factor NNT-1 - useful for treatment of neurological and  
 PT immunological diseases or inflammation, also as vaccine adjuvant  
 PT  
 XX Claim 12; Fig 3; 120pp; English.  
 PS  
 XX This is the amino acid sequence of a novel neurotrophic factor,  
 CC designated NNT-1, that is a growth factor for neurons and for B or  
 CC T cells. It was deduced from isolated cDNA (see AAV47510) and  
 CC genomic DNA (see AAV47511) clones. Vectors containing the cDNA or  
 CC genomic DNA and host cells are provided for use in the production  
 CC of NNT-1 polypeptides. These are used to treat: (i) neurological  
 CC or immunological diseases, specifically Alzheimer's, Parkinson's  
 CC or Huntington's diseases, amyotrophic lateral sclerosis,  
 CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and  
 CC degeneration of the neural retina, or conditions characterised by T  
 CC or B cell defects, e.g. common variable immunodeficiency (CVID),  
 CC selective IgA deficiency, hypogammaglobulinaemia and X-linked  
 CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)  
 CC inflammation. NNT-1 is also able to boost immunoreactivity and  
 CC antibody production following vaccination, and, since it inhibits  
 CC tumour necrosis factor production, it may also be useful for  
 CC treating sepsis. In addition, cells that have been engineered to  
 CC express NNT-1 can be implanted, or nucleic acids are delivered in  
 CC gene therapy vectors.  
 XX  
 SQ Sequence 225 AA;

Alignment Scores:  
 Pred. No.: 5,786-93 Length: 225  
 Score: 1226.00 Matches: 225  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 84.03% Indels: 0  
 DB: 19 Gaps: 0

US-09-931-704-1 (1-797) x AAW29715 (1-225)

QY 90 ATGGACCTCGAGCAGGGGACTGCTGGGGATGTTAGCGTCTGTCAGCGGTCTCTGG 149  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 150 CACCTCCCTGCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATC 209  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
 QY 210 CAGAAACCTATGACTTCACTCCCTACCTGAGCAGACCACTCCGACGCTTGGTGGGACC 269  
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 270 TATCTCACTACTCGGCCCCCTTTCAAGCAGCCAGACTTCAACCTCCCGCCCTGGGG 329  
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 QY 330 GCAGAGACTCTGCCCGGGCCACTGTGTACTTGGAGTGTGGCGAAGCCCTCAATGACAA 389  
 Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
 QY 390 CTCGGCTGACCCAGAACCTACAGGGCTACAGCCACTTCTGTGTACTTGGCTGGCCTC 449

Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 450 AACCGTCAGGCTGCCACTGCTGAGCTGCGCCGAGCTGGCCACTTCTTGACACGCTC 509  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 510 CAGGGCTCTGCTGGCAGCATTCGGGCGTGCATGGCAGCTCTGGGTTACCCACTGCCCCAG 569  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
 QY 570 CCGTCTGCTGGGACTGAACCCACTTGGACTCTGCGCCCTGCCACAGTCACTTCTCCAG 629  
 Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
 QY 630 AAGATGGAGCACTTCTGGCTGTGAAGAGCTGCAGACCTGGCTGGCGCTGGGCGCAAG 689  
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 690 GACTTCAACCGCTCAAGAGATGCAGCTCCAGCAGCTGCAGTGCAGTCCCTGCACCTG 749  
 Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
 QY 750 GGGGCTCATGGCTTC 764  
 Db 221 GlyAlaHisGlyPhe 225

RESULT 3  
 AAW56141  
 ID AAW56141 standard; Protein; 225 AA.  
 XX  
 AC AAW56141;  
 DT 13-JUL-1998 (first entry)  
 XX  
 DE Amino acid sequence of human neurotrophic factor NNT-1.  
 KW Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;  
 KW treatment; neurological disease; degeneration; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..27  
 FT Protein /note= "signal peptide"  
 FT Protein 28..225  
 FT Protein /note= "mature protein"  
 XX  
 PN US5741772-A.  
 XX  
 PD 21-APR-1998.  
 XX  
 PF 03-FEB-1997; 97US-0792019.  
 XX  
 PR 03-FEB-1997; 97US-0792019.  
 XX  
 PA (AMGE-) AMGEN INC.  
 XX  
 PI Chang M;  
 XX  
 DR WPI; 1998-260526/23.  
 DR N-PSDB; AAV22652.  
 XX  
 PT Neurotrophic factor NNT-1 polypeptide and related nucleic acids -  
 PT useful for stimulating growth of motor and sympathetic neurons  
 XX  
 PS Claim 1; Fig 3; 41pp; English.  
 XX  
 CC The present sequence represents a human neurotrophic factor, designated  
 CC NNT-1, which is capable of stimulating growth of motor or sympathetic  
 CC neurons. The NNT-1 protein is useful in the treatment of neurological  
 CC diseases characterised by the degeneration and death of particular  
 CC classes of neurons. These diseases specifically include Parkinson's

CC disease, amyotrophic lateral sclerosis (ALS), Alzheimers's disease,  
 CC stroke and various degenerative disorders affecting vision.  
 XX  
 SQ Sequence 225 AA;

Alignment Scores:  
 Pred. No.: 5,78e-93 Length: 225  
 Score: 1226.00 Matches: 225  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 84.09% Indels: 0  
 DB: 19 Gaps: 0

US-09-931-704-1 (1-797) x AAW56141 (1-225)

QY 90 ATGGACCTCCGAGCGGAGCTCGTGGGGGATGTTAGCGTGCTGTGCAGCGTGCTCTGG 149  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 150 CACCTCCCTGCGAGTCCAGCTCTCAATCCACAGGGGACCCAGGGCCTGGCCCTCCATC 209  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40  
 QY 210 CAGAAACCTATGACCTCACCGCTACCTGGAGCACCACCTCCGAGCTTGGCTGGGACC 269  
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 270 TATCTGAACCTACCTGGGCCCCCTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 329  
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 QY 330 GCAGAGACTTGGCCAGGCCACTTGTGACTTGGAGGTGTGGCAAGCTCAATGACAAA 389  
 Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
 QY 390 CTGGCGCTGAGCCAGACTAGAGCCCTACAGCCACCTTCTGTACTTGGTGGCCCTC 449  
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 450 AACCGTCAAGCTGCGACCTGCTGAGCTGGCGCGCAGCCTGGCCCACTTTCACACCGCTC 509  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 510 CAGGCTGCTGGGCGAGCATTTGGGGCTCATGGCGCTGTGGGTACCCACTGCCCCAG 569  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
 QY 570 CCGTGTGCTGGGACTGAACCCACTGGACTCTGGCCCTGGCCACAGTCACTTCTCCAG 629  
 Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
 QY 630 AAGATGGACGACTTCTGGCTGCTGAAGGAGCTGCAGACTGGCTGGCGCTGGGCCAAG 689  
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 690 GACTTCAACCGCTCAAGAGATGAGCTCCAGAGCTCCAGAGCTGAGTCACTGACCTGACCTG 749  
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaAlaValThrLeuHisLeu 220  
 QY 750 GGGGCTCATGGCTTC 764  
 Db 221 GlyAlaHisGlyPhe 225

RESULT 4  
 AAW94466  
 ID AAW94466 standard; Protein; 225 AA.  
 XX  
 AC AAW94466;  
 XX  
 DT 22-APR-1999 (first entry)  
 XX  
 DE Human cardiotrophin-like cytokine protein.  
 XX

KW Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;  
 KW CLC; IL-6; diagnosis; detection; immune system-related disorder;  
 KW cancer; cardiac disorder; heart failure; hypertension; cancer;  
 KW autoimmune disorder; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..27 /label= signal  
 FT Protein 28..225 /label= Cardiotrophin-like\_cytokine  
 FT Domain 74..79 /label= CD-I  
 FT /note= "conserved domain"  
 FT Domain 150..156 /label= CD-II  
 FT /note= "conserved domain"  
 FT Domain 194..198 /label= CD-III  
 FT /note= "conserved domain"  
 XX  
 PN WO9900415-A1.  
 XX  
 PD 07-JAN-1999.  
 XX  
 PF 29-JUN-1998; 98WO-US13129.  
 XX  
 PR 30-JUN-1997; 97US-0051311.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Shi Y;  
 XX  
 DR WPI: 1999-095678/08.  
 DR N-PSDB; AAX16161.  
 XX  
 PT New isolated cardiotrophin-like cytokine nucleic acid - used to  
 PT develop products for treating cardiac and immune system disorders,  
 PT e.g. heart failure, hypertension, cancers, autoimmune disorders and  
 PT infections  
 XX  
 PS Claim 1; Fig 1; 103pp; English.  
 XX  
 CC The present invention relates to a novel cardiotrophin-like cytokine  
 CC (CLC) protein which is a member of the interleukin 6 (IL-6) cytokine  
 CC family. The present sequence represents the human CLC protein. The  
 CC present invention also describes screening methods for identifying  
 CC agonists and antagonists of CLC activity, as well as methods for  
 CC detecting cardiac and immune system-related disorders and  
 CC therapeutic methods for treating cardiac and immune system-related  
 CC disorders, e.g heart failure, hypertension, cancers, autoimmune  
 CC disorders and infections.  
 XX  
 SQ Sequence 225 AA;

Alignment Scores:  
 Pred. No.: 5,78e-93 Length: 225  
 Score: 1226.00 Matches: 225  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 84.09% Indels: 0  
 DB: 20 Gaps: 0

US-09-931-704-1 (1-797) x AAW94466 (1-225)

QY 90 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCGTGCTGTGCAGCGTGCTCTGG 149  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 150 CACCTCCCTGCGAGTCCAGCTCTCAATCCACAGGGGACCCAGGGCCTGGCCCTCCATC 209  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 210 CAGAAACCTATGACCTCACCGCTACCTGAGCACCACCACTCCGAGCTTGCTGGGACC 269  
DB 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 329  
DB 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80  
QY 330 GCAGAGACTGCCCCAGGGCCACTTGTGACTTGGAGGTGTGGCGAGCTCAATGACAAA 389  
DB 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
QY 390 CTGCGGCTGACCCAGAACCTACGAGCGCTACAGCCACTTCTGTACTTGGTGGCTTC 449  
DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
QY 450 AACCTCAGGCTGCCACTGCTGAGTGGCGCCGAGCTGCGCCACTTTCGACCCGCTTC 509  
DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
QY 510 CAGGGCTGCTGGGAGCATTCGGGGCTCATGGCGCTCATGGCGCTACCCACTGCCCGAG 569  
DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
QY 570 CGCTGCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCGCCAGCTGCTGCTG 629  
DB 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
QY 630 AAGATGGACGACTTCTGGCTGTGAGGAGCTGAGAGCTGCGAGCTGCTGGGCTACCCACTGCCCGAG 689  
DB 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
QY 690 GACTTCAACCGGCTCAAGAGAGATGAGCTGAGCTGCGAGCTGCGAGCTGCGAGCTG 749  
DB 201 AspPheAsnArgLeuLysLysMetGlnProAlaAlaValThrLeuHisLeu 220  
QY 750 GGGGCTCATGGCTTC 764  
DB 221 GlyAlaHisGlyPhe 225

RESULT 5  
AAY87813  
ID AAY87813 standard; Protein; 225 AA.  
AC AAY87813;  
XX 24-AUG-2000 (first entry)  
DT Human NNT-1 protein.  
DE NNT-1; human; neurotrophic factor; neurotrophic; neuroprotective; treatment;  
KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;  
KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;  
KW amphotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
KW Huntington's disease; peripheral neuropathy; neural retina degeneration;  
KW retinopathy; immune disorder; hematopoietic disorder.  
XX  
OS Homo sapiens.  
XX  
FN US6054294-A.  
XX  
PD 25-APR-2000.  
XX  
PF 12-DEC-1997; 97US-0988819.  
XX  
XX 03-FEB-1997; 97US-0792019.  
PR  
PA (AMGE-) AMGEN INC.  
XX  
PI Chang M;  
XX  
DB WPI; 2000-338492/29.

DR N-PSDB; AAA39481.  
XX New nucleic acids encoding neurotrophic factors useful for stimulating  
PT growth of motor or sympathetic neurons for treating neuron cell damage  
PT  
XX  
XX Claim 1c; Fig 3; 42pp; English.  
XX  
CC This invention describes a novel nucleic acid molecule (I) encoding a  
CC novel neurotrophic factor (NNT-1) (II) which has neurotrophic,  
CC neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and  
CC polypharmacological activity. (I) is useful for producing NNT-1  
CC polypeptides which are useful for treating patients in whom various  
CC cells of the central, autonomic, or peripheral nervous system have  
CC degenerated and/or have been damaged by congenital disease, trauma,  
CC mechanical damage, surgery, stroke, ischemia, infection, metabolic  
CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1  
CC proteins are used to treat diseases like Alzheimer's, Parkinson's,  
CC amphotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's  
CC disease, peripheral neuropathy induced by diabetes or other metabolic  
CC disorders, and/or dystrophies or degeneration of the neural retina such  
CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of  
CC night blindness, progressive cone-rod degeneration, immune disorders and  
CC hematopoietic disorders. (II) is effective in treating neurological  
CC conditions and promotes neuron regeneration. Neural functions are  
CC effectively restored in patients suffering from various neurological  
CC disorders. This sequence represents the human NNT-1 protein described in  
CC the method of the invention.  
XX  
SQ Sequence 225 AA;  
Alignment Scores:  
Pred. No.: 5,78e-93 Length: 225  
Score: 1226.00 Matches: 225  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 84.09% Indels: 0  
DB: 21 Gaps: 0  
US-09-931-704-1 (1-797) x AAY87813 (1-225)

QY 90 ATGACCTCCGAGAGCGGAGCTGCTGGGGGATGTAGCGTGTGCGAGCTGCTCTGG 149  
DB 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
QY 150 CACCTCCCTGCGAGTCCAGCTCTCAATCGCACAGAGGAGCCAGGGCTCGGCCCTCCATC 209  
DB 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40  
QY 210 CAGAAACCTATGACCTCACCGCTACCTGGAGCACCACCACTCCGAGCTTGGCTGGGACC 269  
DB 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 329  
DB 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80  
QY 330 GCAGAGACTGCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAGCTCAATGACAAA 389  
DB 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
QY 390 CTGCGGCTGACCCAGAACCTACGAGCGCTACAGCCACTTCTGTGTTACTTGGCTGGCTTC 449  
DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
QY 450 AACCTCAGGCTGCCACTGCTGAGTGGCGCCGAGCTGCGCCACTTTCGACCCGCTTC 509  
DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
QY 510 CAGGGCTGCTGGGAGCATTCGGGGCTCATGGCGCTCATGGCGCTACCCACTGCCCGAG 569  
DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160

QY 570 CCGCTGCTGGGACTGAACCCACTTGGACTCTCGCCCTGCCAGTGAAGTCTCTCCAG 629  
 Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
 QY 630 AAGATGACGACTTCTCGCTGCTGAAGGAGTGTCAGACTCGCTGTGCGCTCGGCCAAG 689  
 Db 181 LysMetAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 690 GACTTCAACCGCTCAGAGAAGATCGAGCTCCAGAGTGCAGTCCACTCGACCTG 749  
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
 QY 750 GGGGCTCATGGCTTC 764  
 Db 221 GlyAlaHisGlyPhe 225  
 RESULT 6  
 ID AAG63543 standard; Protein; 225 AA.  
 AC AAG63543;  
 XX  
 DT 15-OCT-2001 (first entry)  
 XX  
 DE Amino acid sequence of a human NNT-1 protein.  
 XX  
 KW NNT-1; CLF-1; sCNTFRalpha; nervous system; neuron; nervous system;  
 KW neuro-muscular function; tumour; immune system; haematopoietic system;  
 KW reproductive system; liver; skeletal muscle; neurodegenerative disease;  
 KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;  
 KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;  
 KW blastocyst implantation; thrombosis; retinal disease;  
 KW retinal pigmentosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155172-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 26-JAN-2001; 2001WO-FR00253.  
 XX  
 PR 27-JAN-2000; 2000FR-0001035.  
 PR 12-OCT-2000; 2000FR-0013089.  
 XX  
 PA (FABR ) FABRE MEDICAMENT SA PIERRE.  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;  
 DR WPI; 2001-488773/53.  
 DR N-PSDB; AAH74484.  
 XX  
 PT A complex comprising a NNT-1 protein and a CLF-1 and/or sCNTFRalpha  
 PT protein useful to treat neurodegenerative disease including Parkinson's  
 PT and Huntington's, obesity and cancer -  
 XX  
 PS Claim 2; Page 58; 67pp; French.  
 XX  
 CC The present sequence represents a human NNT-1 protein. The specification  
 CC describes a complex comprising a NNT-1 protein and a CLF-1 and/or  
 CC sCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate  
 CC activity of the sCNTFRalpha/gp130/LiFrbeta receptor complex, or to  
 CC induce phosphorylation of the tyrosine of gp130 and LiFrbeta,  
 CC particularly where cells expressing the receptor complex are in the  
 CC central or peripheral nervous system, in neurons implicated in  
 CC neuro-muscular function or in skeletal muscle. The complex or  
 CC antibodies are also used to decrease the survival, growth or  
 CC proliferation of tumour cells or to facilitate the proliferation and/or  
 CC inhibit differentiation of cells stocks. The complex is also used to  
 CC modulate activity of the gp130/LiFrbeta receptor or cells expressing  
 CC that receptor, particularly those cells implicated in the immune,  
 CC haematopoietic, nervous or reproductive system, the liver or skeletal

CC muscle. Molecules of the invention may be used to prevent or treat  
 CC neurodegenerative diseases including amyotrophic lateral sclerosis,  
 CC Parkinson's and Huntington's disease, to repair or regenerate nervous  
 CC or muscular tissue or to maintain muscular mass in paralysis patients.  
 CC They may also be used to treat cancer, obesity and associated diseases,  
 CC and to improve fertility, particularly to avoid endometriosis and/or  
 CC assist blastocyst implantation, thrombosis, or retinal disease,  
 CC particular retinal pigmentosis.  
 XX  
 SQ Sequence 225 AA;  
 Alignment Scores:  
 Pred. No.: 5,78e-93 Length: 225  
 Score: 1226.00 Matches: 225  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 84.09% Indels: 0  
 DB: 22 Gaps: 0

US-09-931-704-1 (1-797) x AAG63543 (1-225)

QY 90 ATGGACCTCCGAGAGGAGTCTGTGGGGATGTTAGCGTGTGTCGACGGTCTCTGG 149  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 150 CACCTCCCTGCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATC 209  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40  
 QY 210 CAGAAAACCTATGACTACCCGCTACCTGGAGCACAACTCCGCGAGCTGGTGGGACC 269  
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 270 TATCTGAACCTACTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTCGGG 329  
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 QY 330 GCAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGGAGAGCCCTCAATGACAAA 389  
 Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
 QY 390 CTGGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACTTGGTGGGCTC 449  
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 450 AACCGTCAGGCTGCCACTGCTGAGCTGCGCGCAGCTGCGCCACTTCTGCACAGCTC 509  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 510 CAGGGCTGCTGGGCGAGCATTTGCGGCGTGCATGCGAGCTCTGGGCTACCCACTGCCAG 569  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
 QY 570 CGCTGCTCGGGAGTCAACCCACTTGGAGTCTCTGGGCTGCGCCAGTGTCTCTCCAG 629  
 Db 161 ProLeuProGlyThrGluProThrTrpTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
 QY 630 AAGATGACGACTTCTGCTGCTGCTGAAGGAGTGCAGACTGCGGTGTGGCTCGGCCAAG 689  
 Db 181 LysMetAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 690 GACTTCAACCGCTCAGAGAAGATGTCAGCTCCAGAGTGCAGTCCACTCGACCTG 749  
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
 QY 750 GGGGCTCATGGCTTC 764  
 Db 221 GlyAlaHisGlyPhe 225  
 RESULT 7  
 ID AAU78176 standard; Protein; 225 AA.  
 XX



AC AAU78176;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Human novel neurotrophic factor NNT1.  
XX  
KW Human, NNT1; neurotrophic factor; IGF-related disease;  
KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;  
KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;  
KW vascular restenosis; rheumatoid arthritis; psoriatic arthritis;  
KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;  
KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;  
KW inflammatory bowel disease; transplant rejection; reproductive disorder;  
KW graft versus host disease; infertility; miscarriage; preterm labour.  
XX  
OS Homo sapiens.  
XX  
XX WO200215977-A2.  
XX  
XX 28-FEB-2002.  
XX  
XX 17-AUG-2001; 2001WO-US25906.  
XX  
XX 18-AUG-2000; 2000US-226436P.  
XX  
XX 16-AUG-2001; 2001US-0931704.  
XX  
XX (AMGE-) AMGEN INC.  
XX  
XX Senaldi G;  
XX  
XX WPI; 2002-280867/32.  
XX  
XX N-PSDB; ABK11647.  
XX  
XX Treating Immunoglobulin E-related disease, modulating IgE levels in a  
XX patient, preventing IGE-related disease and treating allergic diseases,  
XX involves administering NNT-1 inhibitor to a patient  
XX  
XX Claim 2; Fig 3; 63pp; English.  
XX  
XX The invention relates to treating Immunoglobulin E (IgE)-related disease,  
XX modulating IgE levels in a patient, preventing an IGE-related disease,  
XX and treating allergic diseases, comprising administering a  
XX therapeutically effective amount of novel neurotrophic factor (NNT)-1  
XX inhibitor to a patient. Also included are a method of diagnosing an  
XX IGE-related disease or susceptibility to an IGE-related disease, by  
XX determining the presence or amount of expression of an NNT1 polypeptide  
XX encoded by a NNT1 nucleotide sequence, its fragment or naturally  
XX occurring variant, and diagnosing an IGE-related disease or  
XX susceptibility of an IGE-related disease based on the presence or amount  
XX of expression of the polypeptide and a pharmaceutical composition for use  
XX in treating IGE-related disease, comprising the NNT1 inhibitor.  
XX The NNT1 inhibitor is useful for preventing and treating IGE-related  
XX disease, modulating IgE levels, and treating allergic diseases e.g.  
XX Type I allergic disease, allergic rhinitis, eczema, dermatitis,  
XX pollinosis, asthma, immune diseases and disorders, diseases involving  
XX abnormal cell proliferation including cancer, arteriosclerosis and  
XX vascular restenosis, diseases and conditions relating to dysfunction of  
XX immune system including rheumatoid arthritis, psoriatic arthritis,  
XX inflammatory arthritis, osteoarthritis, inflammatory joint disease,  
XX autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory  
XX bowel disease, transplant rejection, and graft versus host disease, and  
XX reproductive diseases and disorders including infertility, miscarriage,  
XX preterm labour and delivery, and endometriosis. The present sequence  
XX represents human NNT1.  
XX  
XX Sequence 225 AA;  
SQ

Alignment Scores:  
Pred. No.: 5,78e-93 Length: 225  
Score: 1226.00 Matches: 225  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 84.09% Indels: 0

DB: 23 Gaps: 0  
US-09-931-704-1 (1-797) x AAU78176 (1-225)  
QY 90 ATGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCTGCTGTGACGGTGTCTTGG 149  
DB 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
QY 150 CACCTCCCTGAGTGGCCAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGGCCCTCCATC 209  
DB 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40  
QY 210 CAGAAAACCTATGACCTCACCGCTACCTGAGCACCACCACTCCGAGCTTGGCTGGACC 269  
DB 41 GlnIysThrTyAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
QY 270 TATCTGAACCTACCTGGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 329  
DB 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
QY 330 GCAGAGACTGCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCTCAATGACAAA 389  
DB 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspIys 100  
QY 390 CTGGGCTGACCCAGAACTACAGGCTTACAGCCACTCTCTGTGTACTTGGCTGGCTC 449  
DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
QY 450 AACCGTCAAGCTGCCACTGCTGAGCTGGCGCGCAGCTGGCCACTTCTGCACCGACCTC 509  
DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
QY 510 CAGGCGCTGCTGGGCGCAGCATTTGCGGGCTCATGGAGCTCTGGGCTACCCACTGCCCCAG 569  
DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
QY 570 CCGTGTCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCCAAGTACTTCTCCAG 629  
DB 161 ProLeuProGlyThrGluProThrTrpProGlyProAlaHisSerAspPheLeuGln 180  
QY 630 AAGTGGACGACTCTGCTGCTGAGGAGCTGAGAGCTGGCTGGCTGGCTGGCTGGCTGG 689  
DB 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaIys 200  
QY 690 CACTTCAACCGCTCAAGAAAGATGCAGCTCCAGCAGCTGCAGTGCAGCTGCAGCTGC 749  
DB 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaAlaValThrLeuHisLeu 220  
QY 750 GGGGCTCATGGCTTC 764  
DB 221 GlyAlaHisGlyPhe 225  
RESULT 8  
ABBI1896  
ID ABBI1896 standard; peptide; 321 AA.  
XX  
AC ABBI1896;  
XX  
DT 11-JAN-2002 (first entry)  
XX  
DE Human cardiotrophin-like cytokine homologue, SEQ ID NO:2266.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibitor; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;

KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnerary; antiulcer.

OS Homo sapiens.

PN WO200157188-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US03800.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457740/49.

XX N-PSDB; ABA09140.

XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -

XX Claim 20; Page 273; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.

XX SQ Sequence 321 AA;

Alignment Scores:

Pred. No.:	7,65e-93	Length:	321
Score:	1225.00	Matches:	232
Percent Similarity:	91.02%	Conservative:	1
Best Local Similarity:	90.62%	Mismatches:	11
Query Match:	84.02%	Indels:	12

DB: 22 Gaps: 3

US-09-931-704-1 (1-797) x ABB11896 (1-321)

Qy 33 CCCACTCCGACAGCTCCGGGAGA---GGAGCCGACACC---GGCCGGCCACCCCGCCAGC 86  
 Db 66 ProHisProProProSerProArgTrpGlyGlnThrProGluGlyLeuProAlaAlaSer 85

Qy 87 CCCATGGAC-----CTCCGACGAGGGGACTCTGTGG 116  
 Db 86 ProCysGlyProGlyProArgSerCysPheSerSerIleLeuProThrGlyAspSerTrp 105

Qy 117 GGGATGTTAGCGTGTGTGCACGGTCTCTGGCACCTCCCTGCAGTGCAGCTCTCAAT 176  
 Db 106 GlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAsn 125

Qy 177 GGCACAGGGGACCCAGGGCTGGCCCTCATCCAGAAAACCTATGACCTCACCCGCTAC 236  
 Db 126 ArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyr 145

Qy 237 CTGGAGCACCAACTCCGACGCTTGGCTGGACCTATCTGAACCTACCTGGGCCCTCTTTC 296  
 Db 146 LeuGluHisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPhe 165

Qy 297 AACGAGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCACTGTT 356  
 Db 166 AsnGluProAspPheAsnProProArgLeuGlyAlaGluThrLeuProAlaThrVal 185

Qy 357 GACTTGGAGGTGGGGAAGCCCTCAATGACAACTGGGCTGACCCAGACTAGGAGCC 416  
 Db 186 AspLeuGluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAla 205

Qy 417 TACAGCCACTTCTGTGTACTTGGCTGGCTCAACCGTCAGGCTGCCACTGTGAGCTG 476  
 Db 206 TyrSerHisLeuLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeu 225

Qy 477 GCGCCGAGCTGGCCCACTTCTGCACAGCTCCAGGGCTGTGGGAGCAGATTGGGGC 536  
 Db 226 ArgArgSerLeuAlaHisPheCysThrSerLeuGlnGlyLeuGlySerIleAlaGly 245

Qy 537 GTCATGGAGCTTGGGCTACCCACTGCCCGCCAGCCGCTGCTGGAGTGAACCCACTGG 596  
 Db 246 ValMetAlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrp 265

Qy 597 ACTCTCTGGCCCTGCCACAGTGTCTCTCCAGAAATGACCACTTCTGGCTGCTGAAG 656  
 Db 266 ThrProGlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLys 285

Qy 657 GAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGACTTCAACCGGCTCAAGAAAGATG 716  
 Db 286 GluLeuGlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMet 305

Qy 717 GAGCTTCAGCAGCTGAGTCACTGACCTGGGCTCATGGCTTC 764

Db 306 GlnProProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 321

RESULT 9

AAW79399

ID AAW79399 standard; Protein; 321 AA.

XX AC AAW79399;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3045.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;

XX KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX XX

PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US04098.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 XX  
 DR N-PSDB; AAK52532.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 PS Claim 20; Page 237; 6221pp; English.  
 XX  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAK78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAK80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 XX  
 SQ Sequence 321 AA;

Alignment Scores:  
 Pred. No.: Length: 321  
 Score: 1225.00 Matches: 232  
 Percent Similarity: 91.02% Conservative: 1  
 Best Local Similarity: 90.62% Mismatches: 11  
 Query Match: 84.02% Indels: 12  
 DB: 22 Gaps: 3

US-09-931-704-1 (1-797) x AAM79399 (1-321)

QY 33 CCCACTCCGCGAGCTCCGGAGA---GGACGCGCACCC---GGCGCGCCAGCCCCAGC 86  
 Db 66 ProHisProProSerProArgTrpGlyGlnThrProGluGlyLeuProAlaAlaSer 85  
 QY 87 CCCATGGAC-----CTCGAGCAGGGGACTCGTGG 116  
 Db 86 ProCysGlyProGlyProArgSerCysPheSerSerIleLeuProThrGlyAspSerTrp 105  
 QY 117 GGGATTTAGCTGCTGTGACGGTGCTGTGGACCTCCCTGACGTGCGCAGCTCTCAAT 176  
 Db 106 GlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAsn 125  
 QY 177 CGCACAGGGGACCCAGGCGCTGCCCTCCATCCAGAAACCTATGACCTCACCCCGCTAC 236  
 Db 126 ArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyr 145  
 QY 237 CTGGAGCACCAACTCCGCGACTTTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTC 296

Db 146 LeuGluHisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPhe 165  
 QY 237 AACGAGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTCCCGAGGCGCACTGT 356  
 Db 166 AsnGluProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrVal 185  
 QY 357 GACTTGGAGGTGTGGCGAAGCCTCAATGACAACTGCGGCTGACCCAGAACCTACGAGGCC 416  
 Db 186 AspLeuGluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAla 205  
 QY 417 TACAGCCACCTTCTGTGTATTCTGCGTGGCTCAACCGTCAAGCTGAGCTGCTGAGCTG 476  
 Db 206 TyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeu 225  
 QY 477 GCGCGCAGCCTGGCCGCTTCTGTCACCGCTCCAGGGCTGCTGGCGAGCATTTGGCGGC 536  
 Db 226 ArgArgSerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGly 245  
 QY 537 GTCTGGCAGCTCTGGGCTTACCCACTGCCCCAGCGCTGCTGGGACTGAACCCACTTGG 596  
 Db 246 ValMetAlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrp 265  
 QY 597 ACTCTGGCCCTGCCACAGTACTTCTCCAGAGATGGACGACTTCTGGCTGCTGAAG 656  
 Db 266 ThrProGlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLys 285  
 QY 657 GAGCTGCAGACCTGGCTGTGGCGCTCGCGCAAGGACTTCAACCGCTCAAGAAGAGATG 716  
 Db 286 GluLeuGlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMet 305  
 QY 717 CAGCCTCCAGCAGCTGCAGTACCCCTGACCTGGGCGGCTCATGGCTTC 764  
 Db 306 GlnProProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 321  
 RESULT 10  
 AAM78415  
 ID AAM78415 standard; Protein; 260 AA.  
 XX AAM78415;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human protein SEQ ID NO 1077.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 OS Homo sapiens.  
 XX  
 FN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US04098.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
 XX

DR WPI; 2001-476283/51.  
 DR N-PSDB; AAK51548.  
 XX Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 PS Claim 20; Page 3306; 6221pp; English.  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAW80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 XX

SQ Sequence 260 AA;

Alignment Scores:  
 Pred. No.: 2,74e-92 Length: 260  
 Score: 1218.00 Matches: 231  
 Percent Similarity: 90.62% Conservative: 1  
 Best Local Similarity: 90.23% Mismatches: 12  
 Query Match: 83.54% Indels: 12  
 DB: 22 Gaps: 3

US-09-931-704-1 (1-797) x AAW78415 (1-260)

QY 33 CCCACTCGCAGCCTCGGGGAGA---GGAGCCGACCC---GGCGGCGCCAGCCCGCAGC 86  
 DB 5 ProHisProProSerProArgTrpGlyGlnThrProGluGlyLeuProAlaAlaSer 24  
 QY 87 CCCATGGAC-----CTCCGAGCGGGGACTCGTGG 116  
 DB 25 ProCysGlyProGlyProArgSerCysPheSerSerIleLeuProThrGlyAspSerTrp 44  
 QY 117 GGGATGTTAGCGTGTCTGACGGTGTCTGGACCTCTCCCTGCGAGTGCAGCTCTCAAT 176  
 DB 45 GlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAsn 64  
 QY 177 CGCAGGGGACCCAGGCGCTGGCCCTCCATCCAGAAACCTATGACCTCAGCCGCTAC 236  
 DB 65 ArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyrAspProThrArgTrp 84  
 QY 237 CTGGAGCACCAACTCCGAGCTTGGCTGGACCTATCTGAACCTACTGGGCGCCCTTTC 296  
 DB 85 LeuGluHisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTrpLeuGlyProProPhe 104  
 QY 297 AACGAGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGCGCACTGTT 356  
 DB 105 AsnGluProAspPheAsnProProArgLeuGlyAlaGluThrLeuProAlaThrVal 124  
 QY 357 GACTTGGAGGTGGCGAGCTCAATGACAACTCCGCTGACCCAGCACTACGAGGCC 416  
 DB 125 AspLeuGluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTrpGluAla 144  
 QY 417 TACAGCACCTTCTGTGTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTG 476  
 DB 145 TyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeu 164  
 QY 477 CGCGCAGCTGGCCCACTTCTGACCAAGCTCCAGGGGCTGCTGGGCGAGCAATGGCGGC 536  
 DB 165 ArgArgSerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGly 184  
 QY 537 GTCATGGAGCTCTGGGCTACCACTGCCCCAGCGCTGCTGGGAGCTCAACCACTTGG 596

DB 185 ValMetAlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrp 204  
 QY 597 ACTCTGGCCCTGCCACAGTGTCTCCAGAGATGGACGACTTCTGGCTGCTGAAG 656  
 DB 205 ThrProGlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLys 224  
 QY 657 GAGCTGCAGACCTGGCTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAAGATG 716  
 DB 225 GluLeuGlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMet 244  
 QY 717 CAGCTCCAGAGCTGCAGTACCTCCCTGCAGCTGGGGCTCATGGCTTC 764  
 DB 245 GlnProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 260  
 RESULT 11  
 AAE00828  
 ID AAE00828 standard; Protein; 223 AA.  
 XX  
 AC AAE00828;  
 DT 02-JUL-2001 (first entry)  
 XX  
 DE Human cardiotrophin-like cytokine (CLC) protein.  
 XX  
 KW Human; biologically active complex; haemopoietin receptor; NR6;  
 KW cardiotrophin-like cytokine; CLC; therapy; prophylaxis; proliferation;  
 KW differentiation; cell survival; neurotrophic activity.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..27 /label= Signal\_peptide  
 FT Protein 28..223 /label= Human\_mature\_CLC\_protein  
 FT /note= "Cardiotrophin-like cytokine"  
 XX  
 PN WO200127157-A1.  
 XX  
 PD 19-APR-2001.  
 XX  
 PF 06-OCT-2000; 2000WO-AU01216.  
 XX  
 PR 08-OCT-1999; 99AU-0003327.  
 PR 12-MAY-2000; 2000AU-0007489.  
 XX  
 PA (AMRA-) AMRAD OPERATIONS PTY LTD.  
 XX  
 PI Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;  
 PI Nakata Y, Hasegawa M;  
 XX  
 DR WPI; 2001-281978/29.  
 DR N-PSDB; AAD04201.  
 XX  
 PT New biologically active complex comprising NR6 and  
 PT cardiotrophin-like cytokine, for facilitating proliferation,  
 PT differentiation and/or survival of a cell -  
 XX  
 PS Claim 32; Page 114-115; 123pp; English.  
 CC The present invention relates to a biologically active complex comprising  
 CC a haemopoietin receptor, NR6 and cardiotrophin-like cytokine (CLC).  
 CC The complex is useful in the manufacture of a medicament for the  
 CC treatment and/or prophylaxis of a subject, as it is involved in  
 CC facilitating proliferation, differentiation and/or survival of a cell.  
 CC The complex or its components have neurotrophic activity. The present  
 CC sequence is human cardiotrophin-like cytokine (CLC) protein.  
 XX  
 SQ Sequence 223 AA;  
 Alignment Scores:  
 Pred. No.: 5,64e-92 Length: 223  
 Score: 1214.00 Matches: 223

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 83.26% Indels: 0  
 DB: 22 Gaps: 0

US-09-931-704-1 (1-797) x AAE00828 (1-223)

QY 90 ATGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCGTGTGCTGACGGTCTCTGG 149  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuGlyValLeuTrp 20  
 QY 150 CACCTCCCTGCGAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGGCCCTCCATC 209  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40  
 QY 210 CAGAAACCTATGACCTACCGCTACCTGGAGCACCACTCCGAGCTGGCTGGGACC 269  
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 329  
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProhrgLeuGly 80  
 QY 330 CGAGAGACTGCCCAGGCGCACTGTGACTTGGAGGTGTGCGAAGGCTCAATCAACAAA 389  
 Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
 QY 390 CTGGCGCTGACCCAGAACCTACAGCGCTACAGCCACCTCTGTGTACTTGGTGGCCTC 449  
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuGlyTyrLeuArgGlyLeu 120  
 QY 450 AACCGCTCAGGCTGCCACTGCTGAGTGGCGCGCAGCTTGGCCACTTTCGACACGACCTC 509  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 510 CAGGCGCTGCTGGGAGCATTCCGGCGCTCATGCGAGCTCTGGGTACCATGCCCCAG 569  
 Db 141 GlnGlyLeuLeuGlySerileAlaGlyValMetAlaLeuGlyTyrProLeuProGln 160  
 QY 570 CGCGTCCCTGGGAGTGAACCTTGGACTCTGGCTGGCCCTGCGCCAGTGCCTCCAG 629  
 Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
 QY 630 AAGATGGAGCACTTCTGGCTGTGAGGAGCTGCAGACTGGCTGTGGCGCTCGGCCAAG 689  
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 690 GACTTCAACCGCTCAAGAGAGATGCGAGCTTCCAGCGCTGAGTGCAGTCCCTGACCTG 749  
 Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
 QY 750 GGGGCTCAT 758  
 Db 221 GlyAlaHis 223

RESULT 12  
 ID AAW29716  
 ID AAW29716 standard; Protein; 225 AA.

XX AAW29716;

XX 09-NOV-1998 (first entry)

XX Mouse neurotrophic factor NNT-1.

XX NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
 KW peripheral neuropathy; dystrophy; neural retina degeneration;  
 KW common variable immunodeficiency; CVID; selective IgA deficiency;  
 KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;  
 KW therapy.

Mus sp.  
 OS  
 XX Key  
 FH Peptide  
 FT Protein  
 FT  
 FT  
 XX  
 PN  
 XX  
 PD  
 XX  
 PF  
 PF  
 PR  
 PR  
 XX  
 PA  
 XX

WO9833922-A1.

06-AUG-1998.

02-FEB-1998; 98WO-US02363.

30-JAN-1998; 98US-0016534.

03-FEB-1997; 97US-0792019.

(AMGE-) AMGEN INC.

Chang M, Elliot GS, Sarmiento U, Senaldi G;

WPI: 1998-437475/37.

N-PSDB; AAV47512.

Newly isolated nucleic acid encoding human or murine neurotrophic factor NNT-1 - useful for treatment of neurological and immunological diseases or inflammation, also as vaccine adjuvant

Claim 13; Fig 5; 120pp; English.

This is the amino acid sequence of a murine neurotrophic factor, designated NNT-1, that is a growth factor for neurons and for B or T cells. It was deduced from isolated NNT-1 cDNA (see AAV47512). Human NNT-1 (see AAW29715) is also provided. Vectors and host cells for use in the production of human murine recombinant NNT-1 polypeptides. These are used to treat: (i) neurological or immunological diseases, specifically Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and degeneration of the neural retina, or conditions characterised by T or B cell defects, e.g. common variable immunodeficiency (CVID), selective IgA deficiency, hypogammaglobulinaemia and X-linked agammaglobulinaemia (claimed), but many others disclosed; and (ii) inflammation. NNT-1 is also able to boost immunoreactivity and antibody production following vaccination, and, since it inhibits tumour necrosis factor production, it may also be useful for treating sepsis. In addition, cells that have been engineered to express NNT-1 can be implanted, or nucleic acids are delivered in gene therapy vectors.

Sequence 225 AA;

Alignment Scores:  
 Pred. No.: 3,06e-90 Length: 225  
 Score: 1193.00 Matches: 218  
 Percent Similarity: 98.22% Conservative: 3  
 Best Local Similarity: 96.89% Mismatches: 4  
 Query Match: 81.82% Indels: 0  
 DB: 19 Gaps: 0

US-09-931-704-1 (1-797) x AAW29716 (1-225)

QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCGTGTGCTGACGGTCTCTGG 149  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuGlyValLeuTrp 20  
 QY 150 CACCTCCCTGCGAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGGCCCTCCATC 209  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40  
 QY 210 CAGAAACCTATGACCTACCGCTACCTGGAGCACCACTCCGAGCTGGCTGGGACC 269  
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 329  
 Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 QY 330 GCAGAGACTGCCCAGGCGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAAA 389  
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100  
 QY 390 CTGGGCTGACCCAGAACCTACGAGCCCTACAGCCACCTCTCTGTGTACTTGGTGGCCCTC 449  
 Db 101 LeuArgLeuThrGlnAsnTyrGlnAlaTyrSerHisLeuLeuGlyLeuArgGlyLeu 120  
 QY 450 AACCGTCAAGCTGCTGAGCTGCGCGCCGAGCCTGCGCCACCTTCTGCACCCAGCCTC 509  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 510 CAGGCGCTGCGGAGCATTTGGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCCAG 569  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160  
 QY 570 CCGTCTGCGGAGCTGAACCCACTTGGACTCTCTGGCCCTGCGCCACAGTCACTTCTCCAG 629  
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180  
 QY 630 AAGATGAGCACTTCTGCTGTGAAGAGCTGCAGACTGCTGGCTGTGGCGCTCGGCCAAG 689  
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 690 GACTTCAACCGGCTCAAGAGAGATGAGCTCCAGAGCTGAGTCACTTCTCCAGCTG 749  
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
 QY 750 GGGGCTCATGGCTTC 764  
 Db 221 GluAlaHisGlyPhe 225

RESULT 13  
 AAW56142  
 ID AAW56142 standard; Protein; 225 AA.  
 AC AAW56142;  
 XX  
 DT 13-JUL-1998 (first entry)  
 XX  
 DE Amino acid sequence of murine neurotrophic factor NNT-1.  
 XX Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;  
 KW treatment; neurological disease; degeneration; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.  
 XX  
 OS Mus sp.  
 XX  
 FH Key  
 FT Peptide 1..27  
 FT Protein /note= "signal peptide"  
 FT /note= "mature peptide"  
 XX  
 PN US5741772-A.  
 XX  
 PD 21-APR-1998.  
 XX  
 PF 03-FEB-1997; 97US-0792019.  
 XX  
 PR 03-FEB-1997; 97US-0792019.  
 XX  
 PA (AMGE-) AMGEN INC.  
 XX  
 PI Chang M;  
 XX  
 DR WPI; 1998-260526/23.  
 DR N-PSDB; AAW22654.

XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -  
 PT useful for stimulating growth of motor and sympathetic neurons  
 XX Claim 2; Fig 5; 41pp; English.  
 XX The present sequence represents a murine neurotrophic factor, designated  
 CC NNT-1, which is capable of stimulating growth of motor or sympathetic  
 CC neurons. The NNT-1 protein is useful in the treatment of neurological  
 CC diseases characterised by the degeneration and death of particular  
 CC classes of neurons. These diseases specifically include Parkinson's  
 CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,  
 CC stroke and various degenerative disorders affecting vision.  
 XX Sequence 225 AA;  
 SQ

Alignment Scores:  
 Pred. No.: 3,06e-90 Length: 225  
 Score: 1193.00 Matches: 218  
 Percent Similarity: 98.22% Conservative: 3  
 Best Local Similarity: 96.89% Mismatches: 4  
 Query Match: 81.82% Indels: 0  
 DB: 19 Gaps: 0

US-09-931-704-1 (1-797) x AAW56142 (1-225)

QY 90 ATGGACTCCGAGCAGGGACTCGTGGGGATGTTAGCGTCCCTGTGCACGGTCTCTGG 149  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 150 CACCTCCCTGCACTGCCAGCTCTCAATCGCACAGGGAGCCAGGGCTGGCCCTCCATC 209  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40  
 QY 210 CAGAAACCTATGACCTCACCGCTACCTGGAGCACCACCTCCGACCTGGCTGGGACC 269  
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 270 TATCTGAACCTACCTGGGCGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGG 329  
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 QY 330 GCAGAGACTTGGCCAGGCGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAAA 389  
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100  
 QY 390 CTGGGCTGACCCAGAACCTACGAGGCTACAGCCACCTCTCTGTGTACTTGGTGGCCCTC 449  
 Db 101 LeuArgLeuThrGlnAsnTyrGlnAlaTyrSerHisLeuLeuGlyLeuArgGlyLeu 120  
 QY 450 AACCGTCAAGCTGCTGAGCTGCGCGCCGAGCCTGCGCCACCTTCTGCACCCAGCCTC 509  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 510 CAGGCGCTGCGGAGCATTTGGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCCAG 569  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160  
 QY 570 CCGTCTGCGGAGCTGAACCCACTTGGACTCTCTGGCCCTGCGCCACAGTCACTTCTCCAG 629  
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180  
 QY 630 AAGATGAGCACTTCTGCTGTGAAGAGCTGCAGACTGCTGGCTGTGGCGCTCGGCCAAG 689  
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 690 GACTTCAACCGGCTCAAGAGAGATGAGCTCCAGAGCTGAGTCACTTCTCCAGCTG 749  
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
 QY 750 GGGGCTCATGGCTTC 764  
 Db 221 GluAlaHisGlyPhe 225

RESULT 14  
 ID AAY87814 standard; Protein; 225 AA.  
 AC AAY87814;  
 DT 24-AUG-2000 (first entry)  
 DE Murine NNT-1 protein.  
 KW NNT-1; neurotrophic factor; neurotropic; neuroprotective; treatment;  
 KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;  
 KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; murine;  
 KW Huntington's disease; peripheral neuropathy; neural retina degeneration;  
 KW retinopathy; immune disorder; hematopoietic disorder.  
 OS Mus sp.  
 PN US6054294-A.  
 PD 25-APR-2000.  
 PF 12-DEC-1997; 9TUS-0988819.  
 PR 03-FEB-1997; 9TUS-0792019.  
 PA (AMGE-) AMGEN INC.  
 PI Chang M;  
 DR WPI; 2000-338492/29.  
 DR N-PSDB; AAA39483.  
 XX New nucleic acids encoding neurotrophic factors useful for stimulating  
 PT growth of motor or sympathetic neurons for treating neuron cell damage  
 PT -  
 PS Claim 2b; Fig 5; 42pp; English.  
 XX This invention describes a novel nucleic acid molecule (I) encoding a  
 CC neurotrophic factor (NNT-1) (II) which has neurotropic,  
 CC neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and  
 CC ophthalmological activity. (I) is useful for producing NNT-1  
 CC polypeptides which are useful for treating patients in whom various  
 CC cells of the central, autonomic, or peripheral nervous system have  
 CC degenerated and/or have been damaged by congenital disease, trauma,  
 CC mechanical damage, surgery, stroke, ischemia, infection, metabolic  
 CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1  
 CC proteins are used to treat diseases like Alzheimer's, Parkinson's,  
 CC amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's  
 CC disorders, and/or dystrophies or degeneration of the neural retina such  
 CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of  
 CC night blindness, progressive cone-rod degeneration, immune disorders and  
 CC hematopoietic disorders. (I) is effective in treating neurological  
 CC conditions and promotes neuron regeneration. Neural functions are  
 CC effectively restored in patients suffering from various neurological  
 CC disorders. This sequence represents the murine NNT-1 protein described in  
 CC the method of the invention.  
 XX  
 SQ Sequence 225 AA;  
 Alignment Scores:  
 Pred. No.: 3,06e-90 Length: 225  
 Score: 1193.00 Matches: 218  
 Percent Similarity: 98.22% Conservative: 3  
 Best Local Similarity: 96.89% Mismatches: 4  
 Query Match: 81.82% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-931-704-1 (1-797) x AAY87814 (1-225)

QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCGTCTGTGCACGGTGTCTCG 149  
 |||||  
 DB 1 MetAspLeuArgAlaIlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 |||||  
 QY 150 CACCTCCCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATC 209  
 |||||  
 DB 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40  
 |||||  
 QY 210 CAGAAACCTATGACCTCACCGCTACCTGGAGACCACTCCGAGCTTGGCTGGGACC 269  
 |||||  
 DB 41 GlnYsThrTyAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 |||||  
 QY 270 TATCTGAACCTACTCGGGGCCCCCTTTCAACAGCAGCAGACTTCAACCCCTCCCGCTGGGG 329  
 |||||  
 DB 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 |||||  
 QY 330 CGAGAGACTCTGCCCGAGGCCACTGTTGACTTGGAGGTGGCGAGCCCTCAATGACAAA 389  
 |||||  
 DB 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100  
 |||||  
 QY 390 CTGCGGCTGACCCAGAACCTACGAGGCTACAGCCACTTCTGTGTTACTTGGTGGGCTC 449  
 |||||  
 DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 |||||  
 QY 450 AACCGTCAAGGCTGCCACTGCTGAGCTGGCGCCGAGCCTGGCCACTTCTGCACCGACCTC 509  
 |||||  
 DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
 |||||  
 QY 510 CAGGGCTGCTGGGAGCATTCGGGGCGTCATGGGAGCTCTGGGTACCCACTGCCCGCCAG 559  
 |||||  
 DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160  
 |||||  
 QY 570 CGCTGCTGGGAGTGAACCCACTTGGACTCTGCTGGCCCTGCCACAGTGCAGTCTCTCCAG 629  
 |||||  
 DB 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180  
 |||||  
 QY 630 AAGATGGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAG 689  
 |||||  
 DB 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 |||||  
 QY 690 GACTTCAACCGGCTCAAGAGAGATGAGCCTCCAGAGCTGCAGTGCAGCTGCACCTG 749  
 |||||  
 DB 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
 |||||  
 QY 750 GGGGCTCATGGCTTC 764  
 |||||  
 DB 221 GluAlaHisGlyPhe 225  
 |||||  
 RESULT 15  
 ID AAY878177  
 ID AAY878177 standard; Protein; 225 AA.  
 AC AAY878177;  
 DT 05-JUN-2002 (first entry)  
 DE Mouse novel neurotrophic factor NNT1.  
 KW Mouse; NNT1; neurotrophic factor; IGE-related disease;  
 KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;  
 KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;  
 KW vascular stenosis; rheumatoid arthritis; psoriatic arthritis;  
 KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;  
 KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;  
 KW inflammatory bowel disease; transplant rejection; reproductive disorder;  
 KW graft versus host disease; infertility; miscarriage; preterm labour.  
 OS Mus sp.  
 PN WO200215977-A2.  
 XX 28-FEB-2002.

XX 17-AUG-2001; 2001WO-US25906.  
 XX PF  
 XX PR  
 XX 18-AUG-2000; 2000US-226436P.  
 PR 16-AUG-2001; 2001US-09311704.  
 XX  
 XX (AMGE-) AMGEN INC.  
 PA  
 XX Senaldi G;  
 PI  
 XX WPI; 2002-280867/32.  
 DR N-PSDB; ABK11649.  
 XX  
 XX Treating Immunoglobulin E-related disease, modulating IgE levels in a  
 PT patient, preventing IgE-related disease and treating allergic diseases,  
 PT involves administering NNT-1 inhibitor to a patient -  
 XX  
 XX Claim 2; Fig 5; 63pp; English.  
 PS  
 XX The invention relates to treating Immunoglobulin E (IgE)-related disease,  
 CC modulating IgE levels in a patient, preventing an IgE-related disease,  
 CC and treating allergic diseases, comprising administering a  
 CC therapeutically effective amount of novel neurotrophic factor (NNT)-1  
 CC inhibitor to a patient. Also included are a method of diagnosing an  
 CC IgE-related disease or susceptibility to an IgE-related disease, by  
 CC determining the presence or amount of expression of an NNT1 polypeptide  
 CC encoded by a NNT1 nucleotide sequence, its fragment or naturally  
 CC occurring variant, and diagnosing an IgE-related disease or  
 CC susceptibility of an IgE-related disease based on the presence or amount  
 CC of expression of the polypeptide and a pharmaceutical composition for use  
 CC in treating IgE-related disease, comprising the NNT1 inhibitor.  
 CC The NNT1 inhibitor is useful for preventing and treating IgE-related  
 CC disease, modulating IgE levels, and treating allergic diseases e.g.  
 CC Type I allergic disease, allergic rhinitis, eczema, dermatitis,  
 CC pollinosis, asthma, immune diseases and disorders, diseases involving  
 CC abnormal cell proliferation including cancer, arteriosclerosis and  
 CC vascular restenosis, diseases and conditions relating to dysfunction of  
 CC immune system including rheumatoid arthritis, psoriatic arthritis,  
 CC inflammatory arthritis, osteoarthritis, inflammatory joint disease,  
 CC autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory  
 CC bowel disease, transplant rejection, and graft versus host disease, and  
 CC reproductive diseases and disorders including infertility, miscarriage,  
 CC preterm labour and delivery, and endometriosis. The present sequence  
 CC represents Mouse NNT1.  
 XX  
 SQ Sequence 225 AA;

Alignment Scores:  
 Pred. No.: 3.06e-90 Length: 225  
 Score: 1193.00 Matches: 218  
 Percent Similarity: 98.22% Conservative: 3  
 Best Local Similarity: 96.89% Mismatches: 4  
 Query Match: 81.82% Indels: 0  
 DB: 23 Gaps: 0

US-09-931-704-1 (1-797) x AAU78177 (1-225)

QY 90 ATGACCTCCGAGCGGGGACTCTGGGGGATGTTAGCGTGCCTGTGACGGTGTCTGG 149  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 150 CACCTCCCTGCGAGTGCAGCTCTCAATCGACAGGGGCCAGGGGCTGGCCCTCCATC 209  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40  
 QY 210 CAGAAACCTATGACCTCACCGCTACTGGAGCCACCACTCCGAGCTGTGGTGGACC 269  
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 270 TATCTGACTACCTGGGGCCCCCTTTCACGAGCCAGACTTCAACCTCCCGCTGGGG 329  
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 330 GCAGAGACTCTGCCAGGGCCACTCTTGCATTGGAGGTGTGGCAGAGCTCAATGACAA 389  
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100  
 QY 390 CTGGGCTGACCCAGAACTACAGGCCCTACAGCCACTTCTGTGTACTTGGTGGCTC 449  
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 450 AACCGTCAGGCTGCCACTGTGAGCTGGCGCCGAGCTGGCCCACTTCTGCACCCAGCTC 509  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 510 CAGGCTCTGCTGGGCGAGCATTCGGGGCGTTCATGGCAGCTCTGGGCTACCCACTGCCCCAG 569  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160  
 QY 570 CCGTGCCTGGGACTGAACCCACTTGGACTCTGGGCCCTGGCCACAGTGACTTCTCCAG 629  
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerPheLeuGln 180  
 QY 630 AAGATGGACGACTTCTGGCTGTGAAGGAGCTGCAGACCTGTGGCTGTGGCTCGGCCAAG 689  
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 690 GACTTCAACCGGCTCAAGAAAGATGCGACCTCCAGCAGCTCAGTCACTCCCTGCACCTG 749  
 Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
 QY 750 GGGGCTCATGGCTTC 764  
 Db 221 GluAlaHisGlyPhe 225

Search completed: January 27, 2003, 15:59:28  
 Job time : 37.5304 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 16:20:23 ; Search time 5.7073 Seconds  
(without alignments)  
5635.701 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 1458

Sequence: 1 attaaagtttcgagagcc.....tctctcttcgtccccccc 797

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 244452

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09931704/runat\_27012003\_154129\_3760/app\_query.fasta\_1.7189  
-DB=Published Applications AA -QPMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-MAXLEN=200000000 -USER=US09931704 @CGN 1 1 51 @runat\_27012003\_154129\_3760  
-NCPV=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pbp.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PT\_NEW\_PUB.pbp.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pbp.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pbp.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pbp.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pbp.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PTUS\_PUBCOMB.pbp.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pbp.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pbp.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pbp.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pbp.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pbp.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pbp.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pbp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1226	84.1	225	10	US-09-931-704-2
2	1193	81.8	225	10	US-09-931-704-5
3	885	60.7	164	10	US-09-864-761-40014
C 4	157.5	10.9	529	10	US-09-861-597-2

Alignment Scores:				
Pred. No.:	3.19e-80	Length:	225	
Score:	1226.00	Matches:	225	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	84.09%	Indels:	0	
DB:	10	Gaps:	0	

#### ALIGNMENTS

RESULT 1  
US-09-931-704-2  
; Sequence 2, Application US/09931704  
; Patent No. US20020041873A1  
; GENERAL INFORMATION:  
; APPLICANT: Senaldi, Giorgio  
; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using Inhibitors  
; FILE REFERENCE: A-695  
; CURRENT APPLICATION NUMBER: US/09/931,704  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: US 60/226,436  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 2  
; LENGTH: 225  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-931-704-2

5	151	10.4	27	10	US-09-864-761-40015	Sequence 40015, A
6	150	10.3	274	10	US-09-850-887-4	Sequence 4, Appli
7	141.5	9.7	507	9	US-10-078-547-24	Sequence 24, Appli
C 8	140	9.7	714	10	US-09-861-597-10	Sequence 10, Appli
9	139.5	9.6	503	9	US-10-078-547-2	Sequence 2, Appli
10	139	9.5	566	10	US-09-791-171-70	Sequence 70, Appli
11	138.5	9.5	802	10	US-09-823-240-2	Sequence 2, Appli
12	137.5	9.4	980	9	US-09-908-193-22	Sequence 22, Appli
13	137	9.4	1274	9	US-10-020-215-2	Sequence 2, Appli
14	136.5	9.4	280	10	US-09-835-232-1	Sequence 1, Appli
15	136.5	9.4	1567	9	US-09-835-232-2	Sequence 2, Appli
16	134	9.2	638	9	US-10-001-887-108	Sequence 108, App
17	131.5	9.0	865	9	US-09-957-998A-19	Sequence 19, Appli
18	129.5	8.9	527	9	US-09-854-133-216	Sequence 216, App
19	129.5	8.9	527	10	US-09-738-973-216	Sequence 216, App
20	128	8.8	674	10	US-09-925-299-979	Sequence 979, App
C 21	126.5	8.7	1004	9	US-09-738-626-5676	Sequence 5676, Ap
22	126	8.6	1247	9	US-09-908-193-2	Sequence 2, Appli
23	125.5	8.6	531	10	US-09-925-300-1444	Sequence 1444, Ap
24	125	8.6	1806	10	US-09-919-437-56	Sequence 56, Appli
25	124.5	8.5	171	10	US-09-864-761-34988	Sequence 34988, A
26	124.5	8.5	171	10	US-09-864-761-36844	Sequence 36844, A
C 27	124.5	8.6	606	10	US-09-861-597-4	Sequence 4, Appli
C 28	123.5	8.5	651	10	US-09-861-597-1	Sequence 1, Appli
29	123	8.4	2442	12	US-10-109-886-10	Sequence 10, Appli
30	122.5	8.4	1134	9	US-10-001-873-50	Sequence 50, Appli
C 31	122	8.4	606	10	US-09-861-597-6	Sequence 6, Appli
C 32	122	8.4	606	10	US-09-861-597-8	Sequence 8, Appli
33	121.5	8.3	847	9	US-10-143-133-2	Sequence 2, Appli
34	120.5	8.3	1422	10	US-09-735-933-1	Sequence 1, Appli
35	120.5	8.3	1690	10	US-09-788-043C-5	Sequence 5, Appli
36	120.5	8.3	4440	9	US-10-174-590-525	Sequence 525, App
37	120.5	8.3	4440	9	US-10-176-758-525	Sequence 525, App
38	120.5	8.3	4440	9	US-10-175-737-525	Sequence 525, App
39	120.5	8.3	4440	12	US-10-052-586-525	Sequence 525, App
40	120	8.2	726	10	US-09-770-689A-4	Sequence 4, Appli
41	120	8.2	803	10	US-09-770-689A-2	Sequence 2, Appli
42	120	8.2	881	10	US-09-816-860A-2	Sequence 2, Appli
43	120	8.2	1665	10	US-09-858-664A-2	Sequence 2, Appli
44	120	8.2	2630	9	US-10-077-130-2	Sequence 2, Appli
45	120	8.2	7968	9	US-10-077-130-5	Sequence 5, Appli

US-09-931-704-1 (1-797) x US-09-931-704-2 (1-225)

QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCGTGTGTCACGGTCTCTGG 149  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
QY 150 CACCTCCCTGACGTCCAGCTCTCAATCGACAGGGGACCCAGGGCTTGGCCCTCCATC 209  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
QY 210 CAGAAACCTATGACCTACCCGCTACTGTGAGCACCACCACTCCGAGCTTGGCTGGGACC 269  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
QY 270 TATCTGAACCTACTGGGCCCCCTTTCAACGAGCCAGCTTCAACCTCCCGCTGGGG 329  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProArgLeuGly 80  
QY 330 GCAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAA 389  
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspLys 100  
QY 390 CTGGGCTGACCCAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGGTGGGCTC 449  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
QY 450 AACCGTCAGGCTGCCACTGTGAGTGGCGGCGAGCTGGCGCCACTTCTGACCCAGCCTC 509  
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
QY 510 CAGGGCTGTGGGACGATTCGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAG 569  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
QY 570 CCGCTGCTGGGAGTGAACCCACTTGACTCTGGCCCTGCGCCACAGTACTTCTCCAG 629  
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
QY 630 AAGATGAGCACTCTGGCTGCTGAAGAGCTGAGACCTGGCTGGCGCTGGCGCCAG 689  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
QY 690 GACTTCAACCGGCTCAAGAAGAAGATGCAGCTCCAGCAGCTGCAGCTCACCTGCACCTG 749  
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
QY 750 GGGGCTCATGGCTTC 764  
Db 221 GlyAlaHisGlyPhe 225

## RESULT 2

US-09-931-704-5  
; Sequence 5, Application US/09931704  
; Patent No. US20020041873A1  
; GENERAL INFORMATION:

; APPLICANT: Senaldi, Giorgio  
; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using N  
; FILE REFERENCE: A-695  
; CURRENT APPLICATION NUMBER: US/09/931.704  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: US 60/226,436  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 225  
; TYPE: PRT  
; ORGANISM: Murine

US-09-931-704-5

Alignment Scores:

Pred. No.: 7,14e-78 Length: 225  
Score: 1193.00 Matches: 218  
Percent Similarity: 98.22% Conservative: 3  
Best Local Similarity: 96.89% Mismatches: 4  
Query Match: 81.82% Indels: 0  
DB: 10 Gaps: 0

US-09-931-704-1 (1-797) x US-09-931-704-5 (1-225)

QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCGTGTGTCACGGTCTCTGG 149  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
QY 150 CACCTCCCTGACGTCCAGCTCTCAATCGACAGGGGACCCAGGGCTTGGCCCTCCATC 209  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
QY 210 CAGAAACCTATGACCTACCCGCTACTGTGAGCACCACCACTCCGAGCTTGGCTGGGACC 269  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
QY 270 TATCTGAACCTACTGGGCCCCCTTTCAACGAGCCAGCTTCAACCTCCCGCTGGGG 329  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProArgLeuGly 80  
QY 330 GCAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAA 389  
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100  
QY 390 CTGGGCTGACCCAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGGTGGGCTC 449  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
QY 450 AACCGTCAGGCTGCCACTGTGAGTGGCGGCGAGCTGGCGCCACTTCTGACCCAGCCTC 509  
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
QY 510 CAGGGCTGTGGGACGATTCGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAG 569  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160  
QY 570 CCGCTGCTGGGAGTGAACCCACTTGACTCTGGCCCTGCGCCACAGTACTTCTCCAG 629  
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180  
QY 630 AAGATGAGCACTCTGGCTGCTGAAGAGCTGAGACCTGGCTGGCGCTGGCGCCAG 689  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
QY 690 GACTTCAACCGGCTCAAGAAGAAGATGCAGCTCCAGCAGCTGCAGCTCACCTGCACCTG 749  
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
QY 750 GGGGCTCATGGCTTC 764  
Db 221 GlyAlaHisGlyPhe 225

## RESULT 3

US-09-864-761-40014  
; Sequence 40014, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmex-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04

```
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 40014
/ LENGTH: 164
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC005849.1
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
/ OTHER INFORMATION: EST HUMAN HIT: AI752561.1, EVALUE 3.00e-66
/ OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUE 8.00e-03
US-09-864-761-40014

Alignment Scores:
Pred. No.: 5,84e-56 Length: 164
Score: 885.00 Matches: 163
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.39% Mismatches: 0
Query Match: 60.70% Indels: 0
DB: 10 Gaps: 0

US-09-931-704-1 (1-797) x US-09-864-761-40014 (1-164)
QY 273 CTGAAGTACTGGGCCCCCTTTCAACGAGCAGACTTCAACCTCCCGCCTGGGGCA 332
Db 1 LeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGlyAla 20
QY 333 GAGACTCTCCCGAGGCCACTGTGTACTTGGAGGTGGCGAAGCCTCAATGACAACTG 392
Db 21 LysThrLeuProArgAlaThrValAspLeuGluValTyrArgSerLeuAsnAspLysLeu 40
QY 393 CGGTGACCCAGAACTACAGAGCCTACAGCCACTTCTGTGTTACTTGTGGCTGCTCAAC 452

/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 40014
/ LENGTH: 164
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC005849.1
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
/ OTHER INFORMATION: EST HUMAN HIT: AI752561.1, EVALUE 3.00e-66
/ OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUE 8.00e-03
US-09-864-761-40014

Alignment Scores:
Pred. No.: 0.000438 Length: 529
Score: 157.50 Matches: 91
Percent Similarity: 32.81% Conservative: 14
Best Local Similarity: 28.44% Mismatches: 122
Query Match: 10.89% Indels: 93
DB: 10 Gaps: 15

US-09-931-704-1 (1-797) x US-09-861-597-2 (1-529)
QY 766 CAGAAGCCATGAGCCCCAGGTGCAGGGTGACTGCAGCTGGAGGTGCATCTCTTC 707
Db 66 GlnGlyProSerGlyProGlySerAlaAlaAlaAlaGly-SerGlyGlnGlnG 85
QY 706 TTGAGCCGGTTCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 686
Db 85 yProGlyGlyTyrGlyProArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 105
QY 685 GCCGAGCGCCACAGCCAGGTCTTCAGCTCTTCAGCAGCCAGAGTGTCTCTCTCG 626
```



GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
Corley, Neil C.  
Gorgone, Gina  
TITLE OF INVENTION: THYROID AND PITUITARY MEMBRANE PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/850,887  
FILING DATE: 07-May-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/087,678  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0535 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 274 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: g206712  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-850-887-4

Alignment Scores:  
Pred. No.: 0.00136 Length: 274  
Score: 150.00 Matches: 71  
Percent Similarity: 34.56% Conservative: 4  
Best Local Similarity: 32.72% Mismatches: 79  
Query Match: 10.29% Indels: 63  
DB: 10 Gaps: 14

US-09-931-704-1 (1-797) x US-09-850-887-4 (1-274)

Qy 15 GGAGCGCGGTGCTCCCTCCACT---CCGCCAGCTCCGGGA----- 54  
Db 55 GlyProGlnArgProGlnGlyProPro-ProProGlyGlyProGlnGlnLysPr 74  
Qy 55 -GAGGAGCGGACCGCGCGGCCAGCCAG-----CCCATGACCTCGGACAGG 107  
Db 74 oGlnGlyProProPro-ProGlyGlyProGlnGlnArgProGlnGlyProProPro 94  
Qy 108 GACTCGTGGGGATTTAGCGGTCTGTGCGCGGTCTGCGACCTCCCTGCGAGTGCCA 167  
Db 94 ly-----GlyProGlnGlnGlyProGlnGlyProProProProGlyG 108  
Qy 168 GCTCTCAATCCGAC-----AGGGGACCCAGGCGCTGGCCCCCTCC 206  
Db 108 lyProGlnGlnArgProGlnGlyProProProProGlyGlyProGlnGlnGlyProGlnG 128  
Qy 207 ATCCAGAAAACCTATGACTCTACCGCTACTGAGGACCACTCGGAGCTTGCTGGG 266  
Db 128 lyPro-----ProProProGlyGly-----ProGlnGlnGlyProG 140

Qy 267 ACCTATCTGAACTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTG 326  
Db 140 lnGlyProProProGlyGlyProGlnGlnSerProProGlnGlyProProProPro 160  
Qy 327 GGGGC---AGAGACTCTGCCAGGCGCCACTGTTGACTTTGGAGGTGTGCGGAAGCCTCAAT 383  
Db 160 lyGlyProGlnGlnGlyProGlnGlyProProProProGlyGly-----ProGln- 176  
Qy 384 GACAACTGCGGCTGACCCAGAACTACGAGCGCTACAGCCACTTCTGTGTACTTGGGT 443  
Db 177 -----GlnSerProProGln-----G 182  
Qy 444 GGCTCAACCGTCAGGTCGCCACTGCTGAGCTGGCGCCGAGCTGGCCCACTTCTGCACC 503  
Db 182 lyProProProProGly-----GlyProGlnGlnGlyPro-----GlnG 195  
Qy 504 AGCTCCAGGCGCTGCTGGGAGCATTCGGGCGCTCATGGGAGCTCTGGGCTACCCACTG 563  
Db 195 lyProProProGlyGlyProGlnGlnArgGlnGly-----ProPro 211  
Qy 564 CCCAGCGCTGCTGGGAGCTGAACCCACTTGGACTCTCTGGCCCT 608  
Db 211 roProGlyGly-ProGlnGlnAspProProGlnGlyProProPro 225

RESULT 7  
US-10-078-547-24  
Sequence 24, Application US/10078547  
Publication No. US20020199211A1  
GENERAL INFORMATION:  
APPLICANT: Narayanasamy Ramesh  
APPLICANT: Miguel A. de la Fuente  
APPLICANT: Ines M. Anton  
APPLICANT: Raif S. Geha  
TITLE OF INVENTION: WIP, A WASP-Associated Protein  
FILE REFERENCE: 1242.1022-005  
CURRENT APPLICATION NUMBER: US/10/078,547  
CURRENT FILING DATE: 2002-02-19  
PRIOR APPLICATION NUMBER: 09/599,287  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: PCT/US98/27501  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/101,457  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/068,533  
PRIOR FILING DATE: 1997-12-23  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: Fast-Seq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 507  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Translated WIP ORF No. US20020199211A1 3  
US-10-078-547-24

Alignment Scores:  
Pred. No.: 0.00599 Length: 507  
Score: 141.50 Matches: 83  
Percent Similarity: 35.33% Conservative: 23  
Best Local Similarity: 27.67% Mismatches: 97  
Query Match: 9.71% Indels: 97  
DB: 9 Gaps: 17

US-09-931-704-1 (1-797) x US-10-078-547-24 (1-507)

Qy 24 GCTCGCCCTCCCACTCCGCCAGCTCCGGAGAGGA----- 59  
Db 127 SerArgProProLeuLeuProGlyGlyArgSerThrSerAlaLysProPheSerPro 146  
Qy 60 ---GCCGACCGCGCGG---CCGAGCCCGAGCCCTCCGAGCGGACTCG 113  
Db 147 ProSerGlyProGlyArgPheProValProSerPro---GlyHisArgSerGly 163

```
QY 114 TGGGGGNTGTTAGCGTGTGACGGTGTCTGGCACCTCCCTGCAGTCCAGCTCTC 173
Db 164 -----ProGluProGlnArg 169
QY 174 AATCGCACAGGGACCCAGGCGCT-----GGCCCTCCATCCAGAAACCTATGACCTC 227
Db 170 AsnArgMetProProProArgProAspValGlySerLys--ProAspSerile---Prop 188
QY 228 ACCCGTACTCGAGACACCACTCCGAGCTGGCTGGGACCTATCTGAACCTACTGGGC 287
Db 188 roProValProSerThrProArgPro---ileGlnSerSerLeuHisAsnArgGlySerP 207
QY 288 CCCCCTTTCAACGAGCCAGACTTCAACCTCCCG-----CCTG 326
Db 207 roProValProGlyProArgGlnProSerProGlyProThrProProPheProG 227
QY 327 GGGCGAGAGACTCTGCGCCAGGCGCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGAC 386
Db 227 lyAsnArgGlyThrAlaLeuGly-----GlyGlySerileArgGlnSerProL 243
QY 387 AACTGGCGGTGACCCA---GAATACAGAGCCCTACAGCCACTTCTGTGTACTTGGCT 443
Db 243 euSerSerSerProPheSerAsnArgProProLeuProProThrProSerArgAlaL 263
QY 444 GG-----CCTCAACCGTCAGGC-----TGCCACTGCT 470
Db 263 euAspAspLysProProProProProProValGlyAsnArgProSerileHisArgG 283
QY 471 GAGCTGGCGCGCAGCTGCGCCACTTCTGACACAGCCTCCAGGCGCTCTGGCGCAGCATT 530
Db 283 luAlaValProProProProGlnAsnAsnLysProProValPro-----298
QY 531 GCGGGCTCATGGCAGCTCTGGGTACCCACTGCCCCAGCGCTGCTGGGACTGAACCC 590
Db 299 -----SerThrProArgProSerAlaProHisArg-----ProH 310
QY 591 ACTTG-----GACTCTCGCGCGCTGCCACAGTGTCTCTCCA-----628
Db 310 isLeuArgProProProSerArgProGlyProProProLeuProProSerSerSerg 330
QY 629 -----GAAGATGGACGACTTCTGGCTGTGTGAAGAGAGTGCAGACCTGGCTGTGGCGC 680
Db 330 lyAsnAspGluThrProArgLeu-----337
QY 681 TCGGCCAGGACTTCAACCGCTCAAGAGAGATGACAGCTCCAGCAGCTCAGTCACCC 740
Db 338 -----ProGlnArgAsnLeuSerLeuSerSerSerThrProProLeuP 352
QY 741 CTGCACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCTCCTCTCGCTCCCGCC 796
Db 352 roSerProGlyArgSerGlyProLeuProProProValProSerGluArgProPro 370

RESULT 8
US-09-861-597-10
; Sequence 10, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861,597
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
```

```
; SEQ ID NO 10
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:protein
US-09-861-597-10

Alignment Scores:
Pred. No.: 0.00804 Length: 714
Score: 140.00 Matches: 70
Percent Similarity: 35.51% Conservative: 17
Best Local Similarity: 28.57% Mismatches: 101
Query Match: 9.68% Indels: 57
DB: 10 Gaps: 12
```

US-09-931-704-1 (1-797) x US-09-861-597-10 (1-714)

```
QY 690 CTTTGGCCGAGCCACAGCCAGGCTCTGAGCTCTTTCAGCAGCCAGAGTCTGTCATCT 631
Db 43 ProGlySerAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlyTyrglyProGlyGln 62
QY 630 TCTGAGGAAGTCACTGTGGGCGAGCCAGGAGTCCAAGTGGGTTCAAGTCCAGGCGCG 571
Db 63 GlnGlyProGlyGlyTyrglyProGlyGlnGlnGlyProSerGlyProGlySerAlaAla 82
QY 570 GCTGGGCGAGTGGTAGCCAGAGCTGCCATGACGCCCGCAATGCTGCCAGCAGGCCCT 511
Db 83 AlaAlaAlaAlaAlaAla-----AlaGlyPro 91
QY 510 GGAGCTGTGTGAGAGTGGGCGAGGCTGGCGCAGCTCAGCAGTGGCAGCTGACGGT 451
Db 92 GlyGlyTyrglyProGly-----GlyProGly-----GlnGlnGlyProGlyGly 103
QY 450 TGAGGCCAGCAAGTAACACAGAGGTGTGTAGGCTGTCTGCTAGTCTGCTGG-----400
Db 104 TyrglyProGlyGlnGlnGlyProGlyGlyTyrglyProGlyGlnGlnGlyProSerGly 123
QY 399 -----TCAGCCGCGAGTTTGTCTATTTAGGCTTCCGCCACCT-----364
Db 124 ProGlySerAlaAlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyProGlyGlyTyrgly 143
QY 363 CCAAGTCAACAGTGGCGCTGGCGAGAGTCTGCCCGCCAGCGGGAGGTTGAAGTCTG 304
Db 144 ProGlyGlnGlnGlyProGlyGlyTyrgly-----ProGlyGlnGlnGlyProSer---160
QY 303 GCTCGTTGAAGGGGGCGCCAGGTAGTTCAGATAGTTCGCCAGCCCAAGCTGCGGAGTTGGT 244
Db 161 -----GlyProGlySerAlaAlaAlaAlaAlaAlaAlaAlaGlyProGly 175
QY 243 GCTCCAGGTAGCGGTGAGTCAATAGTTTCTGGATGGAGGGCCAGGCCCTGGGTCCC 184
Db 176 -----GlyTyrglyProGlyGlnGlnGlyProGlyGlyTyrglyProGly-----190
QY 183 CTGTGCGATTGAGAGCTGGCAGCTGCGAGGAGGTGCCAGACCGTGCACAGCAGCCGTA 124
Db 191 -----GlnGlnGlyProSerGlyProGlySerAlaAlaAla 202
QY 123 ACATCCCCCAGAGTCCCTGCTCGGAGGTCCATGGCGCTGGG---GCTGGCGCGCGCG 67
Db 203 AlaAlaAlaAlaAlaAlaGlyProGlyGly-TyrglyProGlyGlnGlnGlyProGlyGln 222
QY 66 GTGCGGC-----TCCTCTCCCGAGGCTGGCGGAGTGGG---AGGGCGAGCGCGG 19
Db 222 TyrglyProGlyGlnGlnGlyProGlyGlyTyrglyProGlyGlnGlnGlyProSerGln 242
QY 18 CTCCGGCGAGCT 6
Db 242 yProGlySerAla 246
```

RESULT 9  
US-10-078-547-2  
; Sequence 2, Application US/10078547

```
; Publication No. US20020199211A1
; GENERAL INFORMATION:
; APPLICANT: Narayanaswamy Ramesh
; APPLICANT: Miguel A. de la Fuente
; APPLICANT: Ines M. Anton
; APPLICANT: Raif S. Geha
; TITLE OF INVENTION: WIP, A WASP-Associated Protein
; FILE REFERENCE: 1242.1022-005
; CURRENT APPLICATION NUMBER: US/10/078,547
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/599,287
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/068,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Human
; US-10-078-547-2

Alignment Scores:
Pred. No.: 0.00831 Length: 503
Score: 139.50 Matches: 80
Percent Similarity: 35.54% Conservative: 22
Best Local Similarity: 27.87% Mismatches: 88
Query Match: 9.57% Indels: 97
DB: 17 Gaps: 17

US-09-931-704-1 (1-797) x US-10-078-547-2 (1-503)
QY 24 GTCGCTCCCTCCCACTCCGCGACCTCCGGGAGAGA----- 59
Db 124 SerArgProLeuLeuProGlyGlyArgSerThrSerAlaLysProPheSerPro 143
QY 60 ---GCCGACCCGCGCGG---CCGAGCCCGCCGACCTCCGAGCAGGGGACTCG 113
Db 144 ProSerGlyProGlyArgPheProValProSerPro---GlyHiArgSerGly 160
QY 114 TGGGGGATTTAGCTGCTGTGCAGCTGCTCTGGCAGCTCCCTCCGAGTCCCTC 173
Db 161 -----ProProGluProGlnArg 166
QY 174 AATCGCAGAGGGACCCAGGGCCT-----GGCCCTCCATCCAGAAACCTATGAC 227
Db 167 AsnArgMetProProArgProAspValGlySerLys--ProAspSerIle---ProP 185
QY 228 ACCCGCTACCTGGAGCACCACCACTCCGAGCTTGGCTGGGAGCCTATCTGAAC 287
Db 185 roProValProSerThrProArgPro---IleGlnSerSerLeuHisAsnArgGlySerP 204
QY 288 CCCCCTTTCAAGAGCCAGACTTCAACCTCCCG-----CCTG 326
Db 204 roProValProGlyGlyProArgGlnProSerProGlyProThrProProPheProG 224
QY 327 GGGGAGAGACTTCCCGAGGCGCACTGTGACTGGAGGTGGCGAGCCCTCAATGAC 386
Db 224 lyAsnArgGlyThrAlaLeuGly-----GlyGlySerIleArgGlnSerProL 240
QY 387 AAATCGCGCTGACCCA---GAATACGAGGCGCTACAGCCACTTCTGTGTTACTTGG 443
Db 240 euSerSerSerProPheSerAsnArgProProLeuProProThrProSerArgAlaL 260
QY 444 GG-----CCTCAACCGCTCAGGC-----TGCCACTGCT 470
Db 260 euAspAspLysProProProProProValGlyAsnArgProSerIleHisArgG 280
QY 471 GAGTGGCGCGGAGCGCTGGCCCACTTTCTGCACCACTCTCCAGGCGCTGCTGGGCA 530

; Sequence 70, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDLING, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-791-171-70

Alignment Scores:
Pred. No.: 0.00938 Length: 666
Score: 139.00 Matches: 76
Percent Similarity: 32.31% Conservative: 19
Best Local Similarity: 25.85% Mismatches: 99
Query Match: 9.53% Indels: 100
DB: 14 Gaps: 14

US-09-931-704-1 (1-797) x US-09-791-171-70 (1-666)
QY 18 GCGCGGCTCGCCCTCCCACTCCGCGACCTCCG-----GGAGAGAGCGCCACCG-- 69
Db 93 AlaAlaSerLysProProThrProProMet--ProIleAlaGlyProGluProAlaProPr 112
QY 70 -----GCCGGCCCGCCCGCCAGCC-----CCATG 92
```

```
Db 112 oLysProProThrProProMetProIleAlaGlyProGluProAlaProProLysProPr 132
Qy 93 GACCTCCGAGAGGAGGACTCGTGGGGATGTTAGCTGCTGTGCAGCGTCTGGCAC 152
Db 132 oThr-ProProMetProIleAlaGlyProAlaProThrProGluSerGlnLeuAlaP 152
Qy 153 CTCCTGTCAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATCCAG 212
Db 152 roProArgProProThrProGlnThr-----ProThrGlyAlaProGlnGlnProG 169
Qy 213 AAAACCTATGACCTCAC-----CGCTACCTCGGAGCACC 248
Db 169 luSerProAlaProHisValProSerHisGlyProHisGlnProArgArgThrAlaProA 189
Qy 249 CTCGCGAGCTTGGCTGGGACTATCTGAATCACTGTGGGCCCCCTTTCAACAGAGCCAGA- 307
Db 189 laProPro-----TpAlaLysMetProIleGlyGluProProProAlaProSerArgP 207
Qy 308 -----CTTCAACCTCCCGCTGGGAGGAGCTCAATGACAACTCGGCTGACC 341
Db 207 roSerAlaSerProAlaGluProProThrArgProAlaProGlnHisSerArgArgAlaA 227
Qy 342 CCAGGCGGCACTTGTGACTTGGAGGTGGCGGAGCTCAATGACAACTCGGCTGACC 401
Db 227 rgArgGlyHisArgTyArgThrAspThrGluArgAsnValGlyLysValAlaThrGlyP 247
Qy 402 CAGAACTACGA-----GGCC 416
Db 247 roSerIleGlnAlaArgLeuArgAlaGluAlaSerGlyAlaGlnLeuAlaProGlyT 267
Qy 417 TACAGCACTTCTGTGTTACTTGGCTGGCTCAACGTCAGGCTGCGCTGCTGAGCTG 476
Db 267 hrGluProSer-----ProAlaProLeuGly----- 275
Qy 477 CGCGGAGCTGGCGCCACTTCTGCACGAGCTCCAGGCGCTGCTGGGAGCAITTCGCGGC 536
Db 276 -----GlnProArgSerTyLeuAlaProProThrArgProAla----- 288
Qy 537 GTCATGGCAGCTCTGGGCTACCACTGCCCGAGCGGCTGCTGGGAGTGAACCCACTGG 596
Db 289 -----ProThrGluProProProSer----- 295
Qy 597 ACTCTGGCCCTGCCACAGTACTTCTCCAGAAGATGAGCAGCTTCTGGCTGTGAAG 656
Db 296 -----ProSerProGln-----ArgAsnSerGlyArgArgAlaGluArgArgV 310
Qy 657 GAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAAGAGATG 716
Db 310 alHisProAspLeuAla-----AlaGlnHisAlaAlaA 321
Qy 717 CAGCCTCCAGAGCTGCAGTCAACCTGCAGCTGGGGC 754
Db 321 laGlnProAspSerIleThrAlaAlaThrThrGlyGly 333
```

## RESULT 11

```
US-09-823-240-2
; Sequence 2, Application US/09823240
; Patent No. US20020048813A1
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; MOTILITY
; FILE REFERENCE: M0656/7064 (HCL)
; CURRENT APPLICATION NUMBER: US/09/823,240
; PRIORITY FILING DATE: 2001-03-30
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-823-240-2
Alignment Scores:
Pred. No.: 0.0105 Length: 802
Score: 138.50 Matches: 71
Percent Similarity: 35.93% Conservative: 12
Best Local Similarity: 30.74% Mismatches: 84
Query Match: 9.50% Indels: 65
DB: 10 Gaps: 12
US-09-931-704-1 (1-797) x US-09-823-240-2 (1-802)
Qy 30 CCTCCCACTCCGCCA-----GCCTCCGGAGAGAGCGCACCCGCGC---CGGCC 77
Db 381 ProProSerProProIleMetIleSerProProGlyLysAlaThrGlyProArgPro 400
Qy 78 AG-----CCCCAGGCCCATGGACCTCCGAGC----- 103
Db 401 ValLeuProValCysValSerSerProValProGlnMetProProSerProThrAlaPro 420
Qy 104 AGGGGACTCGTGGGGATGTTAGCTGCTGTGTCACGGTGTCTGGCACCTCCCTGCAGT 163
Db 421 AsnGlySerLeuAspSerValThrTyProValSer-----ProProProThrSer 437
Qy 164 GCCAGCTCTCAATCGCACAGGGGAGCGGCTGGCCCTCCATCCAGAAAACCTATGA 223
Db 438 GlyProAlaAlaProProProProProProProProProProProProProPro 453
Qy 224 CCTCACCGCTACCTGGAGCACCACTCCGACGCTT-----GGCTGG 265
Db 454 ProProProLeuProProProProProProProProProLeuAlaSerLeuSerHisCysGlySer 473
Qy 266 GACCTATCTGAACCTACCTGGGCGCCCTTT-----CAACGAGCCAGACTT 310
Db 474 GlnAlaSerProProProGlyThrProLeuAlaSerThrProSerSerLysProSerVal 493
Qy 311 CAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTG 370
Db 494 LeuProSerProSerAlaGlyAlaProAlaSerAla----- 505
Qy 371 GCGAAGCTCAATGACAACTCGGCTGACCCAGAACTA----- 409
Db 506 -----GluThrProLeuAsnProGluLeuGlyAspSerSerAlaSerGlu 520
Qy 410 CGAGGCTACAGCCACTTCTGTGTACTTGGTGGCTCAACCGTCCAGGCTGCCACTGC 469
Db 521 ProGlyLeuGlnAlaAlaSerGlnProAlaGluSerProThrProGlnGlyLeuValLeu 540
Qy 470 TGAGCTGGCGCCAGCTGGCCACTTCTGCACCAGCTCCAGGCTGCTGGGAGCAT 529
Db 541 -----GlyProProAlaProProProProProProProProProPro 554
Qy 530 TCGGGCGCTCATGGCAGCTCTG-----GGCTACCCACTGCCCGCGCTGCC 577
Db 554 yProAlaTyAlaSerAlaLeuProProProProGlyProProProProProProLeuPr 574
Qy 578 TGGGACTGAACCCACTTGGACTCTGGCCCT 608
Db 574 oSerThrGlyProProProProProProProProProPro 584
RESULT 12
US-09-908-193-22
; Sequence 22, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
```



APPLICANT: PADIGARU, MURALIDHARA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY  
; FILE REFERENCE: 21402-062  
; CURRENT APPLICATION NUMBER: US/09/908,193  
; CURRENT FILING DATE: 2001-07-18  
; PRIOR APPLICATION NUMBER: 60/220,273  
; PRIOR FILING DATE: 2000-07-24  
; PRIOR APPLICATION NUMBER: 60/221,650  
; PRIOR FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 60/221,233  
; PRIOR FILING DATE: 2000-07-27  
; PRIOR APPLICATION NUMBER: 60/220,912  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/218,875  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: 60/218,870  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: 60/218,901  
; PRIOR FILING DATE: 2000-07-18  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 980  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-908-193-22

## Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
137.50	137.50	980	79
Percent Similarity:	33.00%	Conservative:	21
Best Local Similarity:	26.07%	Mismatches:	94
Query Match:	9.43%	Indels:	109
DB:	9	Gaps:	15

US-09-931-704-1 (1-797) x US-09-908-193-22 (1-980)

QY	120	ATGTTAGCTGCTGTCACGGTCTCTGG-----CACCTCCCTGCAGTGCAGCT	170
Db	700	LeuLeuAlaCysMetCysAlaGlyLeuArgSerProHisArgGluSerLeuProGly	719
QY	171	CTCAATCGACAGCCAGCCAGCCCTGCGCCCTCCATCCAGAAACCTATGACCTCACC	230
Db	720	LeuSerSerThalaThrProGly---AsnProAlaLeu-----TyrSerArgAla	735
QY	231	CGCTACTGAGCACCACCACTCCGACGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCC	290
Db	736	ArgLeu--GlyProSerProProAlaAlaHisGluLeuGluSerLeuValHisProH	755
QY	291	CCTTTCAACGACGACACTTCAACCTCCCGCTGGGGGACAGACTCTGCCAGGGCC	350
Db	755	isProGln-----AspTrpSerProProSerAspValGluAspArgAlaGluValH	773
QY	351	ACTGTTGACTTGAGGTGGCGAGCCCTCAATGACAAACTGCGGCTGACCCAGAACTAC	410
Db	773	isSerLeuMetGlyGlyValSer-----G	782
QY	411	GAGGCTACAGCCACCTTCTGTGTACTGCGTGGCTCAACCGTCA-----	457
Db	782	luGlyArgSerHisSerLysArgLysIleSerTrpIleGlnProSerGlyLeuSerTrpA	802
QY	458	-----GGCTGCCACTGTGAGCTGCGCCGACGCTGCCCA-----	493
Db	802	laGlySerTrpAlaGlyCysGlu-----LeuProGlnAlaGlyProArgProAlaL	819
QY	494	-----CTTTCGACACGCTTCCAGGGCTCTCTGGGAG-----	526
Db	819	euThrArgAlaLeuLeu-----ProProAlaGlyThrGlyGlnThrLeuLeuGlnA	837
QY	526	-----	526
Db	837	laLeuValTyrAspAlaIleLysGlyAsnGlyArgLysLysSerProProAlaCysArgA	857

QY	527	-----CATTGGGGCGTCAATGGCAGCTCTGGGCTACCCA	560
Db	857	snGlnValGluAlaGluValIleValHisSerAspPheSerAlaSerAsnGlyAsnProA	877
QY	561	CTGCCCCAGCGCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCACAGTGAC	620
Db	877	sp-----LeuHisLeuGlnAspLeuGluProGluAspProL	889
QY	621	TTCTCTCCAGAAGATGACGACTTCTGGCT-----GCTGAAGGAGCTGCAG	665
Db	889	euProProGluAlaProAspLeuIleSerGlyValGlyAspProGlyGlnGlyAlaAlaT	909
QY	666	ACTGTGCTGTGGCTGCGCCAGGACTTCAACCGGCTCAAGAA-----	709
Db	909	rLeuAspArgGluLeuGly--GlyCysGluLeuAlaAlaProGlyProAspArgLeuT	928
QY	710	-----GAAGATGCAGCTCCAGCAGCTGCAGTCACCT-----	742
Db	928	hrCysLeuProGluAlaAlaSerAlaSerCysSerTyrProAspLeuGlnProGlyGluV	948
QY	743	-----GCACCTGGGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCTCGCTC	791
Db	948	alleuGluThrProGlyAspSerCysGlnLeuLysSerProCysProLeuGlyAlaL	968
QY	792	CCCCC	796
Db	968	erPro	969

RESULT 13  
US-10-020-215-2  
; Sequence 2, Application US/10020215  
; Publication No. US20030008347A1  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY  
; APPLICANT: PELES, EIOR  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF ALP RELATED DISORDERS  
; FILE REFERENCE: 038602/1290  
; CURRENT APPLICATION NUMBER: US/10/020,215  
; CURRENT FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: 09/095,443  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/049,477  
; PRIOR FILING DATE: 1997-06-11  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1274  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: ALP  
; OTHER INFORMATION: polypeptide sequence  
US-10-020-215-2

Alignment Scores:  
Pred. No.: 0.0143 Length: 1274  
Score: 137.00 Matches: 74  
Percent Similarity: 32.16% Conservative: 17  
Best Local Similarity: 26.15% Mismatches: 103  
Query Match: 9.40% Indels: 89  
DB: 9 Gaps: 12

US-09-931-704-1 (1-797) x US-10-020-215-2 (1-1274)

QY	19	CGGGGCTCGCCCTCCACTCCGACCTCCGGAGAGAGAGCCGACCCCGCGCGCCCA	78
Db	610	ProGlnGlnProLeuProLeuGlnHisProHisLeuPheProGlnAlaProGlyLeu	629
QY	79	GCCCGACGCCCATGGACCT-----CCGAGCAGGAGCTCGTGGGGATCTTAG	126
Db	630	LeuProProGlnSerProTyrProTyrAlaProGlnProGlyValLeuGlyGlnProPr	649
QY	127	CGTGCTGTGCAGGTGTCTCTGGCACCTCCCTCGACGTCCAGCTCTCAATCGCAGGGG	186

Alignment Scores:  
Pred. No.: 0.0125 Length: 280  
Score: 136.50 Matches: 59  
Percent Similarity: 34.55% Conservative: 7  
Best Local Similarity: 30.89% Mismatches: 60  
Query Match: 9.36% Indels: 66  
DB: 10 Gaps: 10

US-09-931-704-1 (1-797) x US-09-835-232-1 (1-280)

649 oProProLeuHisThrGlnLeuTyrPro-----G1 659  
187 ACCAGGGCTGGCCCTC-----CATCCAGAAACCTATGACCTACCGCTACCTGG 240  
659 yProAlaGlnAspProLeuProAlaHisSerGlyAlaLeu---ProPheProSerProG1 678  
241 AGCACCAACTCCGACGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAACG 300  
678 yProProGlnPro-----ProHisProProLeuAlaTy 689  
301 AGCCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGGCCACTTTGACT 360  
689 rGlyProAlaProSerThrArgProMetGlyProGlnAlaAlaPro-----704  
361 TGGAGGTGGCGAAGCTCAATGACAACTGCGGGTGAGCCAGAACTACGAGCGCTACA 420  
705 -----LeuThrIleArgGlyProSerSerAlaGlyGlnSe 716  
421 GCCACCTTCTGTG---TTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGC 477  
716 rThrProSerProHisLeuValProSerProAlaProSer-----729  
478 GCGCAGCCTGGCCCACTTCTGCACACGCTCCAGGCGCTGCTGGGCGAGCANTCGGGCG 537  
730 -ProGlyProGlyProValProProArgProProAlaAlaGluProProCysLeuAr 749  
538 TCATGGAGCTCTGGG-----CTACCCACTGCGCCAGCCGCTGCTGGGACTGNAC 588  
749 gArgGlyAlaAlaAlaAlaAspLeuLeuSerSerProGluSerGlnHisGlyGlyTh 769  
589 CCACCTGGACTCTCGGCGCTGCCACAGTACTTCCT-----625  
769 rGlnSerProGlyGlyGlnProLeuLeuGlnProThrLysValAspAlaAlaGluG1 789  
626 -----CAGAAGATGG 636  
789 yArgArgProGlnAlaLeuArgLeuLeuGluArgAspProTyrGluHisProGluArgLe 809  
637 ACAGCTCTGGCTCTGAA-----GGAGCTGC 663  
809 uArgGlnLeuGlnGlnGluLeuGluAlaPheArgGlyGlnLeuGlyAspValGlyAlaLe 829  
664 AGACCTGGCTGTGGCTGGCGCCAGGACTTCAACCGGCTCAAGAGAGATGACGCTC 723  
829 uAsp-----ThrValTrpArgGluLeuGlnAspAlaGlnGluHisAspAlaArgG1 846  
724 CAGCAGC 730  
846 yArgSer 848

RESULT 14  
US-09-835-232-1  
; Sequence 1, Application US/09835232  
; Patent No. US2002098489A1  
; GENERAL INFORMATION:  
; APPLICANT: Leder, Philip  
; APPLICANT: Leder, Benjamin  
; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: 00383/052002  
; CURRENT APPLICATION NUMBER: US/09/835,232  
; PRIOR FILING DATE: 2001-04-12  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-835-232-1

Alignment Scores:  
Pred. No.: 0.0125 Length: 280  
Score: 136.50 Matches: 59  
Percent Similarity: 34.55% Conservative: 7  
Best Local Similarity: 30.89% Mismatches: 60  
Query Match: 9.36% Indels: 66  
DB: 10 Gaps: 10

US-09-931-704-1 (1-797) x US-09-835-232-1 (1-280)

QY 30 CCTCCACTCCGACGCTCCGGGAGAGAGCGCCAGCCGCGCGCCAGCCAGCCAG--- 85  
Db 126 ProProProProProProProGlyValGlyIleProProProProProProGlyVal 145  
QY 86 -----CCCCATGGACCTCCGACGAGGAGGACTCTGTGGGGATGTTAGCGTGTGTCAC 139  
Db 146 GlyIleProProProProProProProGlyValGlyIleProProProProProPro 165  
QY 140 GGTGCTCTGGCACCTCTCCGACGAGTCCAGCTCTCAATCCACAGGGGACCCAGGGCTGG 199  
Db 166 GlyValGlyIleProPro-----171  
QY 200 CCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTGGAGCA-----244  
Db 172 -----ProProProProProProGlyValGlyIleProProPro 183  
QY 245 ---CCAACCTCCGACGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTCAACGA 301  
Db 184 ProProProProGlyMetGly-----IleProProProProProPro 198  
QY 302 GCCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGGCCACTTGTGACTT 361  
Db 199 GlySerGlyIleProProProProAla-----Leu 208  
QY 362 GGAGCTGTGGCAAGCCTCAATGACAACTGCGGCTGACCCAGAACTA-----CGAGGC 415  
Db 209 ProGlyValAlaIlePro-----ProProProProProProGlyMetGly 223  
QY 416 CTACAGCCACCTTCTGTGTACTTGTGCTGGCTCAACCGTACAGGTGCCACTGCTGAGCT 475  
Db 224 ValProProProAla-----ProProProProGly-----Ala 234  
QY 476 GCGCGCAGCCTGGCGCCACTTCTGCACACGCTCCAGGGCGCTCTGGGCGAGCATTGGCGG 535  
Db 235 GlyIleProProProProLeuLeu-----ProGlySerGlyProProHisSerSer 251  
QY 536 CGTCTATGGCAGCTCTGGGCTACCCACTGCCCA 568  
Db 252 GlnValGlySerSerThrLeuProAlaAlaPro 262

RESULT 15  
US-09-835-232-2  
; Sequence 2, Application US/09835232  
; Patent No. US2002098489A1  
; GENERAL INFORMATION:  
; APPLICANT: Leder, Philip  
; APPLICANT: Leder, Benjamin  
; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: 00383/052002  
; CURRENT APPLICATION NUMBER: US/09/835,232  
; PRIOR FILING DATE: 2001-04-12  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1567  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-835-232-2

Alignment Scores:





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:50:02 ; Search time 8.08533 Seconds  
(without alignments)  
5800.643 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 1458

Sequence: 1 attaaagctgcgcgagcc.....tctctcttcgtccccccc 797

Scoring table: BLOSUM62

          Xgapop 10.0 , Xgapext 0.5  
          Ygapop 10.0 , Ygapext 0.5  
          Fgapop 6.0 , Fgapext 7.0  
          Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlp

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09931704/runat\_27012003\_154127\_3671/app\_query.fasta\_1.7189  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=ral -MINMATCH=0.1 -LOOFCFL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09931704 @CGN 1.1.89 @runat\_27012003\_154127\_3671 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1226	84.1	225	1	US-08-792-019B-2
2	1226	84.1	225	3	US-09-106-182-2
3	1226	84.1	225	3	US-08-988-819-2
4	1226	84.1	225	4	US-09-016-534-2
5	1193	81.8	225	1	US-08-792-019B-5
6	1193	81.8	225	3	US-08-988-819-5
7	1193	81.8	225	4	US-09-016-534-5
8	160	11.0	1064	1	US-08-642-255-62
9	159	10.9	960	4	US-09-219-849-5
10	157.5	10.8	330	1	US-08-642-255-32
11	157.5	10.8	408	1	US-07-609-716-65
12	157.5	10.8	408	4	US-08-475-411A-65

13	157.5	10.8	408	4	US-08-478-029A-65	Sequence 65, Appl
14	157.5	10.9	529	4	US-09-247-806-2	Sequence 2, Appl
15	155.5	10.8	870	2	US-09-010-928B-2	Sequence 2, Appl
16	154	10.7	493	4	US-08-556-978B-59	Sequence 59, Appl
17	152	10.5	595	1	US-08-425-069-4	Sequence 4, Appl
18	152	10.5	595	2	US-08-317-844B-4	Sequence 4, Appl
19	148	10.2	252	1	US-08-642-255-61	Sequence 61, Appl
20	144.5	10.0	745	2	US-09-010-928B-28	Sequence 28, Appl
21	144	10.0	1008	4	US-09-219-849-8	Sequence 8, Appl
22	144	9.9	1057	3	US-08-931-820-4	Sequence 4, Appl
23	144	10.0	1065	1	US-08-642-255-80	Sequence 80, Appl
24	144	10.0	1065	3	US-08-642-246-16	Sequence 16, Appl
25	144	10.0	1065	4	US-09-451-206-16	Sequence 16, Appl
26	144	10.0	1065	5	PCT-US96-06229-16	Sequence 16, Appl
27	141	9.7	1341	3	US-08-963-825-18	Sequence 18, Appl
28	141	9.7	1341	4	US-09-500-811-18	Sequence 18, Appl
29	141	9.7	1341	4	US-09-570-573-18	Sequence 18, Appl
30	141	9.7	1341	4	US-09-548-608-18	Sequence 18, Appl
31	140	9.7	714	4	US-08-556-978B-61	Sequence 61, Appl
32	140	9.7	714	4	US-09-247-806-10	Sequence 10, Appl
33	139	9.5	905	2	US-08-574-959A-9	Sequence 9, Appl
34	139	9.5	905	4	US-09-357-014-9	Sequence 9, Appl
35	138.5	9.6	960	4	US-09-219-849-6	Sequence 6, Appl
36	138	9.5	1078	3	US-08-963-825-21	Sequence 21, Appl
37	138	9.5	1078	4	US-09-500-811-21	Sequence 21, Appl
38	138	9.5	1078	4	US-09-570-573-21	Sequence 21, Appl
39	138	9.5	1078	4	US-09-548-608-21	Sequence 21, Appl
40	137.5	9.4	1461	4	US-09-585-887-9	Sequence 9, Appl
41	137.5	9.4	1461	4	US-09-289-578-9	Sequence 9, Appl
42	137	9.4	357	1	US-07-609-716-66	Sequence 66, Appl
43	137	9.4	357	1	US-08-642-255-33	Sequence 33, Appl
44	137	9.4	357	4	US-08-475-411A-66	Sequence 66, Appl
45	137	9.4	357	4	US-08-478-029A-66	Sequence 66, Appl

#### ALIGNMENTS

#### RESULT 1

US-08-792-019B-2  
; Sequence 2, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUROTROPIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/792,019B  
; FILING DATE: 03-FEB-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-792-019B-2

## Alignment Scores:

Pred. No.: 3,75e-102 Length: 225  
Score: 1226.00 Matches: 225  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 84.09% Indels: 0  
DB: 1 Gaps: 0

US-09-931-704-1 (1-797) x US-08-792-019B-2 (1-225)

QY 90 ATGGACCTCCGAGCAGGGGACTCTGGGGGATGTTAGCGTCTGTGCACGGTGTCTGG 149  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
QY 150 CACTCTCCCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGCCCTCCATC 209  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40  
QY 210 CAGAAACCTATGACCTACCCGCTACCTGGAGCACCAACTCCGAGGCTTGGCTGGGACC 269  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCAGACTTCAACCTCCCGCTGGGG 329  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProArgLeuGly 80  
QY 330 GCAGAGACTCTGCCAGGCCACTGTGACTTGGAGGTGGCGAAGCTCAATGACAAA 389  
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
QY 390 CTGGGCTGACCCAGAACTACGAGGCCTTACGAGCCTCTGTTGTTACTTGGTGGCCTC 449  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
QY 450 AACGCTCAGGCTGCACCTGCTGAGCTGGCGCGAGCTGGCCCACTTCTGCACGACCTC 509  
Db 121 AsnArgGlnAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
QY 510 CAGGGCTCTGGGACAGATTGCGGGGCTCATGCGAGCTGCGAGCTGGGCTACCCACTGCC 569  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
QY 570 CCCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTCCAGCAGTGCACCTGCACCTG 629  
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
QY 630 AAGATGACGACTTCTGGCTGCTGAGGAGCTGAGAGCTGCGAGCTGGCTGGCGCTCCAG 689  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
QY 690 GACTTCAACCGGCTCAAGAGATGAGCTCCAGCAGCTGCGAGTGCACCTGCACCTG 749  
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
QY 750 GGGGCTCATGCTTC 764  
Db 221 GlyAlaHisGlyPhe 225

## RESULT 2

US-09-106-182-2  
; Sequence 2, Application US/09106182  
; Patent No. 6046035

## GENERAL INFORMATION:

; APPLICANT: Shi, Yangu  
; APPLICANT: Ruben, Steve  
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc  
; STREET: 9410 Key West Ave  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: US

; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/106,182  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/051,053  
; FILING DATE: 30-JUN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF385  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-106-182-2

## Alignment Scores:

Pred. No.: 3,75e-102 Length: 225  
Score: 1226.00 Matches: 225  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 84.09% Indels: 0  
DB: 1 Gaps: 0

US-09-931-704-1 (1-797) x US-09-106-182-2 (1-225)

QY 90 ATGGACCTCCGAGCAGGGGACTCTGGGGGATGTTAGCGTCTGTGCACGGTGTCTGG 149  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
QY 150 CACTCTCCCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTGCCCTCCATC 209  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40  
QY 210 CAGAAACCTATGACCTACCCGCTACCTGGAGCACCAACTCCGAGGCTTGGCTGGGACC 269  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCAGACTTCAACCTCCCGCTGGGG 329  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProArgLeuGly 80  
QY 330 GCAGAGACTCTGCCAGGCCACTGTGACTTGGAGGTGGCGAAGCTCAATGACAAA 389  
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
QY 390 CTGGGCTGACCCAGAACTACGAGGCCTTACGAGCCTCTGTTGTTACTTGGTGGCCTC 449  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrProLeuProGln 160  
QY 450 AACGCTCAGGCTGCACCTGCTGAGCTGGCGCGAGCTGGCCCACTTCTGCACGACCTC 509  
Db 121 AsnArgGlnAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
QY 510 CAGGGCTCTGGGACAGATTGCGGGGCTCATGCGAGCTTGGGCTACCCACTGCCCGG 569  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
QY 570 CCGCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTCCAGCAGTGCACCTGCCTCC 629  
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180

QY	630	AGATGAGCAGCTTCGGCTGCTGAAGAGCTCGACACCTGGCTGGCGCTCGCGCAAG	689
Db	181	LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTripleuTrpArgSerAlaLys	200
QY	690	GACTTCAACCGGCTCAAGAGAAAGATGCAGCCTCCAGCAGCTGCAGCTCACCTTGACACCTG	749
Db	201	AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaValThrLeuHisLeu	220
QY	750	GGGGCTCATGCTTC 764	
Db	221	GlyAlaHisGlyPhe 225	
RESULT 3			
US-08-988-819-2			
; Sequence 2, Application US/08988819			
; Patent No. 6054294			
; GENERAL INFORMATION:			
; APPLICANT: CHANG, MING-SHI			
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1			
; NUMBER OF SEQUENCES: 16			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: AMGEN INC.			
; STREET: ONE AMGEN CENTER DRIVE			
; CITY: THOUSAND OAKS			
; STATE: CA			
; COUNTRY: USA			
; ZIP: 91320			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/988,819			
; FILING DATE: 12-DEC-1997			
; CLASSIFICATION: 536			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/792,019			
; FILING DATE: 03-FEB-1997			
; ATTORNEY/AGENT INFORMATION:			
; NAME: COOK, ROBERT R.			
; REGISTRATION NUMBER: 31,602			
; REFERENCE/DOCKET NUMBER: A-442A			
; INFORMATION FOR SEQ ID NO: 2:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 225 amino acids			
; TYPE: amino acid			
; TOPOLOGY: linear			
; MOLECULE TYPE: protein			
US-08-988-819-2			
Alignment Scores:			
Pred. No.: 3,75e-102 Length: 225			
Score: 1226.00 Matches: 225			
Percent Similarity: 100.00% Conservative: 0			
Best Local Similarity: 100.00% Mismatches: 0			
Query Match: 84.09% Indels: 0			
DB: 3 Gaps: 0			
US-09-931-704-1 (1-797) x US-08-988-819-2 (1-225)			
QY	90	ATGGACCTCCGACGAGGGGACTCGTGGGGATCTTAGCGTGCTGTGCGCTGCTCG	149
Db	1	MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp	20
QY	150	CACCTCCCTCGAGTCGCAGCTCTCAATCGCACAGGGGACCCAGGGGCTTGCCCCCTCCATC	209
Db	21	HisLeuProAlaValProAlaLeuAsnArgThr-GlyAspProGlyProGlyProSerIle	40
QY	210	CAGAAACCTATGACCTACCCGCTACCTGGAGACCAACTCCGAGCTTGGTGGGACC	269
Db	41	GlnLysThrTrpAspLeuThrArgTrpLeuGluHisGlnLeuArgSerLeuAlaGlyThr	60

```

; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-016-534-2

```

**Alignment Scores:**

Pred. No.:	3,75e-102	Length:	225
Score:	1226.00	Matches:	225
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	84.09%	Indels:	0
DB:	4	Gaps:	0

US-09-931-704-1 (1-797) x US-09-016-534-2 (1-225)

QY	90	ATGGACCTCCGAGCAGGGGACTCGTGGGGGATCTTACGTGCTGTGACGGTGTCTCGG	149
Db	1	MetAspLeuArgAlaGlyAspSerTrpGlyWetLeuAlaCysLeuCysThrValLeuTrp	20
QY	150	CACCTCCCTCAGTGCAGCTCTCAATCGACACAGGGAGCCACGGGCTCGCCCTCCATC	209
Db	21	HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle	40
QY	210	CAGAAAACTATGACCTCAACCCCTCACTTGGAGCACCAACTCCGACGTGTGGTGGACC	269
Db	41	GlnIysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr	60
QY	270	TACTTGAACCTACCTGGGCCCCCTTTCACAGAGCCAGACTTCAACCCCTCCCGCTGGG	329
Db	61	TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly	80
QY	330	GCAGAGACTCTGCCACGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAA	389
Db	81	AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys	100
QY	390	CTGGCGGTGACCCAGAACTACGAGGGCTACAGCCACTTCTGTGTACTTGGTGGCTC	449
Db	101	LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu	120
QY	450	AACCGTCAAGCTGCCTGCTGAGCTGCGCGCAGCGCTGCGCCACATCTTCGACACAGCCTC	509
Db	121	AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu	140
QY	510	CAGGCGCTGCTGGCAGCATTGGGGGCTCATGGAGCTCTGGGCTTACCCACTGCCCCAG	569
Db	141	GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaLeuGlyTyrProLeuProGln	160
QY	570	CCGTGCTCGGGAGTGAACCCACTTGGACPTCTGGCGCTGCCACACAGTACACTTCCTCCAG	629
Db	161	ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln	180
QY	630	AAGATGACGACTTCTGGCTGTGAAGGAGCTCGACACTGGCTGTGGCGCTCGGCCAAG	689
Db	181	LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys	200
QY	690	GACTTTCAACCGGCTCAAGAGAAGATGCAGCCCTCCAGCAGCTGCAGTCAACCTCGACCTG	749
Db	201	AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu	220
QY	750	GGGGCTCATGGCTTC	764
Db	221	GlyAlaHisGlyPhe	225

## RESULT 5

US-08-792-019B-5  
; Sequence 5, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUTROTROPIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESSES:  
;





TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-016-534-5

Alignment Scores:  
Pred. No.: 3,426-99 Length: 225  
Score: 1193.00 Matches: 218  
Percent Similarity: 98.22% Conservative: 3  
Best Local Similarity: 96.88% Mismatches: 4  
Query Match: 81.82% Indels: 0  
DB: 4 Gaps: 0

US-09-931-704-1 (1-797) x US-09-016-534-5 (1-225)

```

QY 90 ATGGACCTCGAGCAGGAGCTGTGGGGATGTTAGCGTGTGCTGCACGGTCTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
QY 150 CACCTCCCTGCAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
QY 210 CAGAAACCTATGACCTACCCCTACCTGGAGCACCAACTCCGACCTGGCTGGACC 269
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGAACCTACCTGGCCGCCCTTTCAACAGGACGACACTCAACCTCCCGCTGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProArgLeuGly 80
QY 330 GCAGAGACTCTGCCAGGGGCACCTTTGACTTGGAGTGTGGGAGCCTCAATGACAAA 389
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100
QY 390 CTGCGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTTACTTGGCTGCTC 449
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACCGTCAGCTGCCTGCTGAGCTGCGCGGAGCTGCGGACCTTCTGACACGCTC 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
QY 510 CAGGGCTGCTGGGAGCATTGGGGCTCATGCAGCTCTGGGCTACCCACTGCCCCAG 569
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
QY 570 CCCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCACAGTGACTTCTCCAG 629
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AGATGACGACTTCTGGCTGTGAAGAGCTGAGAGCTGCGAGCTGGCTGTGGCTCGGCCAAG 689
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 690 GACTTCAACCGGCTCAGAGAAGATCAGCTCCAGCAGCTGAGTCAACCTGCACCTG 749
Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
QY 750 GGGGCTCATGGCTTC 764
Db 221 GluAlaHisGlyPhe 225

```

# RESULT 8

US-08-642-255-62  
Sequence 62, Application US/08642255  
Patent No. 5773249  
GENERAL INFORMATION:  
APPLICANT: CARPELLO, Joseph  
TITLE OF INVENTION: High Molecular Weight Collagen-Like  
TITLE OF INVENTION: Protein Polymers  
NUMBER OF SEQUENCES: 135

CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/642,255  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: ROWLAND, Bertram I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: A5556-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299 FHT UR  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1064 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-642-255-62

Alignment Scores:  
Pred. No.: 2,616-06 Length: 1064  
Score: 160.00 Matches: 80  
Percent Similarity: 37.66% Conservative: 7  
Best Local Similarity: 34.63% Mismatches: 81  
Query Match: 10.97% Indels: 63  
DB: 1 Gaps: 15

US-09-931-704-1 (1-797) x US-08-642-255-62 (1-1064)

```

QY 4 AAAGTCTCGCGAGCCGCGGCTCGCCCTCCACTCCGACGCTCCGGAGAG----- 57
Db 135 ArgGlyAspProGlyProGlyAlaProGlyProAlaGlyProGlySerArgGly 154
QY 58 GAGCGCACCGCGCGCGGCGGAGCCGCGGAGCCGCTCCG---AGCAGGGGACTCGT 114
Db 155 AspProGlyPro-ProGlyAlaProGlyProAlaGlyProGlySerArgGlyAspPr 174
QY 115 GGGGATGTTAGCGTGTGCTGCAGGCTCTGCGCACCTCCCTGCGAGTCCAGCTCTCA 174
Db 174 oGlyProProGlyAlaPro-----GlyProAlaGlyProPro-----G1 187
QY 175 ATCGCACAGGGGACCCAGGGCT-----GGCCCTCCATCCAGA 213
Db 187 YSerArg-GlyAspProGlyProProGlyAlaProGlyProAlaGly-ProProGlySer 206
QY 214 AACCTATGACTCACCCTACTCTGGAGCACCAACTCCGAGCTTGGCTGGGACCTATC 273
Db 207 ArgGlyAsp-ProGlyProPro-----ProGlyAlaProGlyPro----- 218
QY 274 TGAACCTACTGGGCGCCCTTTCAACGAGCCAGACTTCAACCTCCCGGCTGGG----- 328
Db 219 -----AlaGlyProProGlySerArgGly-----AspProGlyProProGlyAlaPr 234
QY 329 -----GGCAGAGACTCTGCCCGGAGCCACTCTTGACTTGGAGGTGTGGCGAAGCC 378
Db 234 oGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGlyAlaProGlyPr 254
QY 379 TCAATGACAAACTGGGCTGAGCCAGAACTACGAGGCTACAGCCACTTCTGTGTACT 438

```

Db 254 o-----AlaGlyProProGlySerArgGly----- 262  
QY 439 TGGGTGGCTCAACCTCAGGCTGCCACTGCTGAGCTGGCGCCAG- 484  
Db 263 ----AspProGlyProProGlyAlaHisGlyProAlaGlyProGlySerArgGlyPr 291  
QY 485 -CCTGCCCACTT---CTGCACAGCCTCCAGGCGCTGCTGGCGCAGCATTTGGCGGGGTCA 540  
Db 281 oAlaGlyProGlyAlaHisGlyProAlaGlyProGlySerArgGlyAlaHis--GlyProAl 300  
QY 541 TGGCAGCTCTGGGCTACCACTGCCAGCGCTGCTGGGACTGAA-----C 588  
Db 300 aGlyProGlyAlaProGlyProAlaGlyProGlySerArgGlyAspProGlyP 320  
QY 589 CCACTTGGAGCTCCTGCCCTGCC 611  
Db 320 roProGlyAlaProGlyProAla 327  
RESULT 9  
US-09-219-849-5  
; Sequence 5, Application US/09219849  
; Patent No. 6150081  
; GENERAL INFORMATION:  
; APPLICANT: VAN HEERDE, GEORGE V.  
; APPLICANT: VAN RIJN, ALEXIS C.  
; APPLICANT: BOWMSTRA, JAN B.  
; APPLICANT: DE WOLF, FREDERIK A.  
; APPLICANT: MOOROEK, ANDREAS  
; APPLICANT: WERTEN, MARC W.T.  
; APPLICANT: WIND, RICHIELE D.  
; APPLICANT: VAN DEN BOSCH, TANJA J.  
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN  
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE  
; FILE REFERENCE: 2728-2  
; CURRENT APPLICATION NUMBER: US/09/219,849  
; CURRENT FILING DATE: 1998-12-23  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; TYPE: PRT  
; LENGTH: 960  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative  
; OTHER INFORMATION: amino acid sequence  
US-09-219-849-5  
Alignment Scores:  
Pred. No.: 3.le-06 Length: 960  
Score: 159.00 Matches: 74  
Percent Similarity: 40.82% Conservative: 6  
Best Local Similarity: 37.76% Mismatches: 87  
Query Match: 10.91% Indels: 29  
DB: 4 Gaps: 11  
US-09-931-704-1 (1-797) x US-09-219-849-5 (1-960)  
QY 46 CCTCCGGGA---GAGGAGCGCCAGCGCGCGCCAGCCAGCCGATGAGCTCCGAG 102  
Db 111 ProGlySerArgAspProGlyProProGlyAlaProGlyProAlaGlyProProGl 130  
QY 103 CAGGGGACTCGTGGGATTTAGCGTGCCTGTGCAGCGTGTCTGGCAGCTCCCTGCAG 162  
Db 130 ySerArgAspProGlyProProGlyAlaPro-----GlyProAlaGlyProProGlySe 148  
QY 163 TGGCAGCTCTCAATCGCACAGGAGCCAGGCGCTGGCGCTCCATCCAGAAACCTATG 222  
Db 148 rArgAspPro-GlyProProGlyAlaProGlyPro-AlaGlyPro-ProGlySerArgAs 167  
QY 223 ACCTACCCGCTACCTGAGCAGCACTCGCAGCTTGGCTGGGACCTATCTGAACTACC 282  
Db 167 pProGlyPro---ProGlyAlaProGlyPro---AlaGlyProProGlySerArgAsPr 185

QY 283 TGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGGGAGAGACTCTGC 342  
Db 185 OGlyProPro-----GlyAlaProGlyProAlaGlyProProGlySerArgAsp----- 201  
QY 343 CAGGGGCACTGTTGACTTGGAGGTGTGGAGAGCTCAATGACAAACTGCGGCTGACCC 402  
Db 202 -ProGlyProProGlyAlaProGlyProAlaGlyPro-----ProGlySerArgAspPr 219  
QY 403 AGAACTACGAGGCTACAGCCACTTCTGTGTTACTTGGTGGCCTCAACCGTCAGGCTG 462  
Db 219 OGlyProProGlyAlaHisGlyProAla-----GlyProGlyAla 233  
QY 463 CCACTGCTGAGCTGCGCCGCGAGCTGCTGCCACTTCTGCACCGCTCCAGGGCTGCTGG 522  
Db 233 aHisGlyProAlaGlyProGlyAla-----HisGlyProAlaGlyProGlySGL 250  
QY 523 GCAGCATTTGGGCGTCATGGCAGCTCTGGGCTACCCAGTCCCGCCAGCGCTCCCTGGGA 582  
Db 250 yAlaHis---GlyProAlaGlyProGlyAlaProGly-ProAlaGlyProProGlyS 269  
QY 583 CTGAA-----CCCACTTGGAGCTCCTGGCGCTGCC 611  
Db 269 erArgAspProGlyProProGlyAlaProGlyProAla 281  
RESULT 10  
US-08-642-255-32  
; Sequence 32, Application US/08642255  
; Patent No. 5773249  
; GENERAL INFORMATION:  
; APPLICANT: CAPPELLO, Joseph  
; APPLICANT: FERRARI, Franco A.  
; TITLE OF INVENTION: High Molecular Weight Collagen-Like  
; TITLE OF INVENTION: Protein Polymers  
; NUMBER OF SEQUENCES: 135  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/642,255  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROWLAND, Bertram I.  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: A55556-3/BIR  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 910 277299 FHT UR  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 330 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-642-255-32  
Alignment Scores:  
Pred. No.: 2.96e-06 Length: 330  
Score: 157.50 Matches: 72  
Percent Similarity: 37.00% Conservative: 2  
Best Local Similarity: 36.00% Mismatches: 90



NUMER OF SEQUENCES: 119  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,411A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/609,716  
FILING DATE: 06-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/269,429  
FILING DATE: 09-NOV-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 408 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-475-411A-65

Alignment Scores:  
Pred. No.: 3,178-06 Length: 408  
Score: 157.50 Matches: 72  
Percent Similarity: 37.00% Conservative: 2  
Best Local Similarity: 36.00% Mismatches: 90  
Query Match: 10.80% Indels: 36  
DB: 4 Gaps: 12

US-09-931-704-1 (1-797) x US-08-475-411A-65 (1-408)  
QY 30 CCTCCACTCCGACCTCCGGAGAGGAGCGCACCGCGCGCCAGCCCGCCCGCC 89  
Db 41 ProProGlyProProGlyProProGlyProProGlyProProGlyProProGly 60  
QY 90 ATGGACCTCGAGC---AGGGACTCTGGGGGATTTAGCTGCT---GTGCAGGTG 143  
Db 60 roGlyProProGlyProAlaGlyProValGlySerProGlyAlaProGlyProProGly 80  
QY 144 CTCTGGCACTCTCCAGTCAGTCCTCAATCGCACAGGAGGACCGCGCGCCCGCC 203  
Db 80 roProGlyProProGlyProProGlyProProGlyProProGlyProProGly 100  
QY 204 TCCATCCAGAAACCTATGACCTCACCGGTACCTGGAGCACCACCACTCCGAGCTTGGCT 263  
Db 100 lyProProGlyProAlaGlyProValGlySerProGlyAlaProGlyPro----- 116  
QY 264 GGGACCTATCTGAACTACTGGCGCCCGCCCTTTCAACGAGCAGACTTCAACCTCCCGC 323

Db 117 -----ProGlyProProGly-----ProProGlyProP 126  
QY 324 CTGGG-----GGCAGAGACTCTGCCAGGCGCACTGTGACTTGGAGGTGGCGAAGC 377  
Db 126 roGlyAlaProProGlyProProGlyProProGlyProProGlyProProGlyPro 146  
QY 378 CTCAATGACAAACTCGCGCTGACCCAGAACACTAGAGGCTACAG---CCACTTCTGTGT 434  
Db 146 ro-----ValGlySerProGlyAlaProGlyProProGlyProProGlyPro- 161  
QY 435 TACTTGGCTGGCCTCAACCGTCAGCTGCCACTGTGAGCTGGCGGCGAGCTGGCCAC 494  
Db 162 -----ProGlyProProGlyAlaProGlyProProGlyPro- 176  
QY 495 TTCTGCACCACTCCAGGCGCTGTGGCGAGCATTTGGCGGCTCATGCGAGCTCTGGGC 554  
Db 177 -----ProGlyProProGlyProAlaGlyPro-ValGlySerProGlyAlaProGly 194  
QY 555 TACCCACTGCCCGCGCTGCTGGGACTGAACCCACTTGGACTCTGGCCT 608  
Db 194 roProGlyProProGlyPro-ProGly-----ProProGlyAlaProGlyPro 209  
RESULT 13  
US-08-478-029A-65  
; Sequence 65, Application US/08478029A  
; Patent No. 6184348  
; GENERAL INFORMATION:  
; APPLICANT: Ferrari, Franco A.  
; TITLE OF INVENTION: Functional Recombinantly Prepared  
; TITLE OF INVENTION: Synthetic Protein Polymer  
; NUMBER OF SEQUENCES: 119  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,029A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/609,716  
; FILING DATE: 06-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/269,429  
; FILING DATE: 09-NOV-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/114,618  
; FILING DATE: 29-OCT-1987  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 06/927,258  
; FILING DATE: 04-NOV-1986  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-781-1989  
; TELEFAX: 415-398-3249  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 408 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

! MOLECULE TYPE: protein  
US-08-478-029A-65

Alignment Scores:  
Pred. No.: 3,17e-06 Length: 408  
Score: 157.50 Matches: 72  
Percent Similarity: 37.00% Conservative: 2  
Best Local Similarity: 36.00% Mismatches: 90  
Query Match: 10.80% Indels: 36  
DB: 4 Gaps: 12

US-09-931-704-1 (1-797) x US-08-478-029A-65 (1-408)

```
QY 30 CTTCCACTCCGAGCTCGGGAGAGGAGCCACCCCGCCGCGCCAGCCCGCC 89
Db 41 ProProGlyProProGly-ProProGlyAlaProGlyProProGlyPro- 60
QY 90 ATGACCTCCGAGC--AGGGGACTCTGGGGGATGTTAGCGTCCT---GTGCACGGTG 143
Db 60 roGlyProProGlyProAlaGlyProValGlySerProGlyAlaProGlyPro- 80
QY 144 CTCTGGCACTCTCCCTGAGTGCAGCTCTCNAATCGCACAGGGGACCCAGGCCCTGGCCCC 203
Db 80 roProGlyProProGlyProProGlyAlaProGlyProProGlyProProG 100
QY 204 TCCATCCAGAAACCTATGACCTCACCCGCTACCTGGAGCACCACTCCGACGCTTGGCT 263
Db 100 lyProProGlyProAlaGlyProValGlySerProGlyAlaProGlyPro- 116
QY 264 GGGACCTATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCC 323
Db 117 -----ProGlyProProGly-----ProProGlyProp 126
QY 324 CTGGG-----GGCAGAGACTCTCCAGGCGCCACTGTTGACTTGGAGGTGGGAGC 377
Db 126 roGlyAlaProGlyProProGlyProProGlyProProGlyProProGlyProAlaGlyP 146
QY 378 CTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCTACAG--CCACTTCTGTGT 434
Db 146 ro-----ValGlySerProGlyAlaProGlyProProGlyProPro- 161
QY 435 TACTTGGTGGCTCAACCGCTAGGCTGCCACTGCTGAGCTGCGCCGAGCTGGCCAC 494
Db 162 -----ProGlyProProGlyAlaProGlyProProGlyPro--ProGlyPro- 176
QY 495 TTCTGACACAGCTCCAGGCGCTGCTGGGAGCATTGGGGGCTCATGGCAGCTTGGGC 554
Db 177 -----ProGlyProProGlyProAlaGlyPro---ValGlySerProGlyAlaProGlyP 194
QY 555 TACCCAGTCCCGCCGCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCT 608
Db 194 roProGlyProProGlyPro-ProGly-----ProProGlyAlaProGlyPro 209
```

## RESULT 14

US-09-247-806-2  
; Sequence 2, Application US/09247806  
; Patent No. 6280747  
; GENERAL INFORMATION:  
; APPLICANT: PHILLIPPE, Michel  
; APPLICANT: GARSON, Jean-Claude  
; APPLICANT: ARRAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; TITLE OF INVENTION: ANALOG  
; FILE REFERENCE: 6388-0365-0  
; CURRENT APPLICATION NUMBER: US/09/247,806  
; CURRENT FILING DATE: 1999-02-11  
; EARLIER APPLICATION NUMBER: FR 98/01614  
; EARLIER FILING DATE: 1998-02-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 529

! TYPE: PRT  
; ORGANISM: Nephila clavipes  
US-09-247-806-2

Alignment Scores:  
Pred. No.: 3,46e-06 Length: 529  
Score: 157.50 Matches: 91  
Percent Similarity: 32.81% Conservative: 14  
Best Local Similarity: 28.44% Mismatches: 122  
Query Match: 10.89% Indels: 93  
DB: 4 Gaps: 15

US-09-931-704-1 (1-797) x US-09-247-806-2 (1-529)

```
QY 766 CAGAAGCATGAGCCCGCCAGGTGACGTGAGCTGGAGGCTGATCTTCTTC 707
Db 66 GlnGlyProSerGlyProGlySerAlaAlaAlaAlaAlaGly-SerGlyGlnGlnG 85
QY 706 TTGAGCCGTTGAAGTCCT-----TG 686
Db 85 yProGlyGlyTyroGlyProArgGlnGlnGlyProGlyGlyTyroGlyGlnGlnGlnG 105
QY 685 GCCGAGCGCCACAGCAGGTCTGAGCTCTTCAGCAGCCAGAGTCGTCATCTTCTGG 626
Db 105 yProSerGlyProGlySerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyG 125
QY 625 AGGAAGTCACTGTGGCGGCGCAGAG-----598
Db 125 yProGlyGlyTyroGlyProGlyGlnGlnGlyProGlyGlyTyroGlyProGlyGlnGlnG 145
QY 597 -----TCCAAGTGGGTTTCACTGTCAGTCCAGGCGCAGC 572
Db 145 yProGlyGlyTyroGlyProGlyGlnGlnGlyProSerGlyProGlySerAlaAlaAla 165
QY 571 GCTGGGCGAGTGGGTAGCCAGAGCTGCATGACGCCCGCATGTGTCGCCAGCA-----517
Db 165 aAlaAlaAlaAlaSerGlyPro-----GlyGlnGlnGlyProGlyGlyTy 180
QY 516 -----GGCCCTGGAGGCTGG-----TGCAGAAAGTGGGCC 488
Db 180 rGlyProGlyGlnGlnGlyProGlyGlyTyroGlyProGlyGlnGlnGlyProSerGlyP 200
QY 487 AGGCTGCGCGCGAGCTCAG-----CAGTGGCAGCTGACGG 452
Db 200 oglySerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyP 220
QY 451 TTGAGCCAGCGCAAGTAACAGAGGTGGCTGATGGCT-----CGTAG 407
Db 220 yTyroGlyProGlyGlnGlnGlyProGlyGlyTyroGlyProGlyGlnGlnGlyLeuSerG 240
QY 406 TTCTGGGTGAGCGCGAGTTTGTTCATTGAGGCTTCGCCACACCTCCAAGTCAACAGTGGCC 347
Db 240 yProGlySerAlaAlaAla-----AlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyP 257
QY 346 CTGGGCGAGAGTCTCTCCCGCCAGCGGGAGGTTGAAGTCTGGCTGTTGAAGGGGGG 287
Db 257 oglyGlyTyroGly-----ProGlyGlnGlnGlyProSer-----G 269
QY 286 CCCAGTAGTTCAGATAGTCCCGCAGCAAGCTCGGAGTTGTTGCTCCAGTAGCGGGTG 227
Db 269 yProGlySerAla-----AlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlyTyroGlyP 288
QY 226 AGGTATAGTGTTCCTGGATGGAGGGGCCAGGCC---CTGGGTCCCTGTGCGATTGAGA 170
Db 288 oglyGlnGlnGlyProGlyGlyTyroGlyProGlyGlnGlnGlyPro-----303
QY 169 GCTGCACTGCGAGGAGTGCAGACACCGTGCACAGGACCGCTAACATCCCCACAG 110
Db 304 -----SerGlyAlaGlySerAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyG 319
QY 109 TCCCTGCTCGAGGTCCATGGGCTGG---GCTGGCGCGCGGCTGGCG---60
Db 319 nGlnGlyLeuGlyGly-TyroGlyProGlyGlnGlnGlyProGlyGlyTyroGlyProGlyG 339
```

QY 407 GTTCTGGGTACGCCAGCTTTGTTCATTAGAGCTTCGCCACACCTCCCAAGTCAACAGTGGC 348  
Db 153 laSerGlyProSerGlySerGlyProSerGlySerArgProSerSerSerGlyProSer 173  
QY 347 CCTGGCAGAGTCTCTGCCCCCAGCGGGAGGTTGAAGTCTGGCTCGTTGAAAAGGGG 288  
Db 173 erGlyThrArgProSerProSerGlyAlaSerGlySerSer-----ProGlyG 189  
QY 287 GC-----CCAGGTAGTTTCAGATAGTCCAGCCCAAGCTGGCGAGTTGGTGTCCAGGTA 234  
Db 189 lyIleAlaProGlyGlySerAsn-----SerGlyGlyAlaGlyValSerGlyA 205  
QY 233 CGGGTGTAGTTCATAGTTTCTGTGATGAGGGGGCCAGCCCTGGCTCCCTGTGGATT 174  
Db 205 laThrGlyGlyProAlaSerSerGly-----SerTyrGlyPro----- 217  
QY 173 GAGAGTGGCACTGCAGGGAGGTGCCAGAGCACCGTGCACAGGCACGCTAACATCCCCA 114  
Db 218 -----GlySerThrGlyGlyThrTyrGlyProSerGlyGly-----SerGluProP 233  
QY 113 CGAGTCCC---CTGCTCGAGGTCCATGG-----GGCTGGGGCTGGGCGCGCGGTGC 63  
Db 233 heGlyProGlyValAlaGlyGlyProTyrSerProGly-GlyAlaGlyProGlyGlyAla 252  
QY 62 GGCTCCTCTCCCGAGGTGGCGGAGTGGAGGGGAGCGCGCGGTCCG 14  
Db 253 GlyGlyAlaTyrGly-ProGlyGlyValGlyThrGlyGlyAlaGlyPro 268

Search completed: January 27, 2003, 16:25:43  
Job time : 14.0853 secs

QY 59 --TCCTCTCCCGAGGTGGCGAGTGGAGGGCGAGCGCGGTCCGGCGAAGCT 6  
Db 339 lngInGlyProGlyGlyTyrGlyProGlySerAlaSerAlaAlaAlaAla 357  
RESULT 15  
US-09-010-928B-2  
; Sequence 2, Application US/09010928B  
; Patent No. 5994099  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V  
; APPLICANT: Hayashi, Cheryl Y  
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA  
; TITLE OF INVENTION: CODING THEREFOR  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
; STREET: 8110 GATEHOUSE RD. SUITE 500E  
; CITY: FALLS CHURCH  
; STATE: VIRGINIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 22042  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/010,928B  
; FILING DATE: 22-JAN-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M  
; REGISTRATION NUMBER: 28977  
; REFERENCE/DOCKET NUMBER: 1447-109P  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 870 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-010-928B-2

Alignment Scores:  
Pred. No.: 6,18e-06 Length: 870  
Score: 155.50 Matches: 76  
Percent Similarity: 40.17% Conservative: 20  
Best Local Similarity: 31.80% Mismatches: 108  
Query Match: 10.75% Indels: 35  
DB: 2 Gaps: 10

US-09-931-704-1 (1-797) x US-09-010-928B-2 (1-870)  
QY 701 CCGGTTGAAGTCTTGGCGAGCGCCACAGCAGTCTCGAGTCTCTCAGCAGCCAGAA 642  
Db 58 ProAsnGluGluPheValHisGluValGlnAspLeuIleGlnMetLeuSerGlnGluGln 77  
QY 641 GTCGTCCATCTTCTGGAGGAAGTCACTGTGGGAGGAGCCAGGA-----GTCCAAAGTGGG 588  
Db 78 Ile-----RanGluValAspThr-SerGlyProGlyGlnTyrTyrArgSerSe 93  
QY 587 TTCAGTCCAGGCGCGGTGGGCGAGTGGGTAGCCAGAGTGGCCATGACCCCGCAAT 528  
Db 93 rSerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyProValValThrGluThrLeuThrVa 113  
QY 527 GCTGCCAGCAGGCGCCCTGGAGGCTGGTGACAGAGTGGCCAGGCTGGCGGCAGCTCAGC 468  
Db 113 i-ThrValGlyGlySerGlyGlyGlyGlnProSerGlyAlaGlyProSerGlyThrGlyG 133  
QY 467 AGTGGCAGCTGACGGTTGAGGCCACGCAAGTAACACAGAGGTGGCTGTAGGCTCGTA 408  
Db 133 lyTyralaProThrGlyTyrAlaProSerGlySerGlyAlaGlyGlyValArgProSerA 153







D	b	809	SerGlyGlySerSerGlyAspSerAlaAlaGlyGlyGlySerSerArgGlyTyr	828
Q	y	1137	TGATAGACACTCATGGTGCCAAAGTCCTCATCTCTCAACAATACACATGCCTCTCTTTCTC	1196
D	b	829	-----:::--SerGluGlySerSerGlyGlyGlyGlyGlySerGly	842
Q	y	1197	TCCCGTCTTGCAGGAGTGTTCCTCTCTCCATCCCTCTGCCTCCCATTCTGGTGTCCC	1256
D	b	843	GlyTyrSerGlyGlySerAlaProProProProProProProProProProProAlaPro	862
Q	y	1257	ACCCTACCCCCCACGCCAAGTGGGGACAGACACTGAGGGGCTGCCAGCTGCTT	1316
D	b	863	AlaProAlaProAlaProSerGlyGly-	872
Q	y	1317	CCCCGTGGGCCCGGGCCGCTCATGCTTCTCGTCCATCTGCCCCACAGGGGACTCGT	1376
D	b	873	-----TyrSerGlyGlySerSerGlySerAlaAlaGlyGlyGlyGlySerSer	890
Q	y	1377	GGGGATGTTTAGCTGCTGCACGGTCTCTGGCACCTCTCCCTGCAGTCCACGCTCTCA	1436
D	b	891	GlyGly-----TyrSerGlyGlySerSerSerGlyGlyGlyGlyGlySerSerGlyGly	908
Q	y	1437	ATFCGACAGGGGACCCAGGGCCGT---GCCCTCCATCCAGAAAACCTATGACTCACCC	1493
D	b	909	TyrSerGlyGlySerAlaAlaLeuProProProProProProProProProProProPr	928
Q	y	1494	GCTACTCGGACCAACTCCGACGCTTGGCTGGGACCTATGTGAGTATCCACGCTAGGA	1553
D	b	928	oProProAlaProAlaProAla-	940
Q	y	1554	ATCTGGGAGTTGGGAGGAGTGAGGAGTTGGGAAAGACAGTCTTAACCGTGAGGGTTC	1613
D	b	940	rSerGlyGlyTyrSerGlyAlaSerSerGlySerAlaAlaGlyGlyGlyGlyGlySe	960
Q	y	1614	TGTTAAATGATGGGTGAGGAGGCTCTTTGGCTCCACCACTCCCTCTGTCTGTCTTA	1673
D	b	960	rSerGlyGlyTyrSer--GlyGlySerAlaAlaProProProProPro-	975
Q	y	1674	TCTCTGCCCCCTCTTAGGTGGCCCCCCCCCACTTCCCATCCCTGGCCCCCAGGACTAG	1733
D	b	976	-----ProProProProProPro-	981
Q	y	1734	GCATGTGGCAGGCTCGCACCCGCTTTGGCCCATTTGCCCACTGGCTGCCAGCCAGCC	1793
D	b	982	-----ProProProProProAlaPro--AlaProAlaProAla	993
Q	y	1794	GCCCGCTCCCTCGGGGCGGGGAAGTCTCTCTGTTTACCCGTGTGTGGTGTCTC	1853
D	b	994	-ProSerSerGlyGlyTyrSerGlyGlySer-----SerGlyGlySe	1007
Q	y	1854	TTGCGCGGGCGGGTTGGTGGGGACAGGG-----GCCCCAC	1892
D	b	1007	rAlaAlaGlyGlyGly-GlyGlySerSerGlyGlyTy-rSerGlyGlySerAlaAlaProP	1027
Q	y	1893	CTCCCATCCCTGCGTTCAGCTCGCTCTGCCCCACAGACTGGGGCCCTGCTCTGGA	1952
D	b	1027	roProProPro-----ProProProProAla-	1036
Q	y	1953	CCAGGGGCTCCCTTCGCTGCTCCATCTCCATCTAGCTGGGCTCTTAGGGGGTCTAT	2012
D	b	1037	-----ProAlaProAlaProAlaProSerSerGlyGlyTyrSerGlyGlySerS	1053
Q	y	2013	GGGGAGGGGACTGTAGGGAACCCAGCAGTAGTGGCAGGGGTTTAGGGTGTGGATGG	2072
D	b	1053	erGlyGlySerAlaAlaGlyGlyGlyGlyGlySerSerGlyGlyTyrSer-----	1069
Q	y	2073	AGGTTATCTGTAGGATTTGGGGTGGTCCAGAGGTTCACAGAGCCCGCAGAGAAG	2132
D	b	1069	-----	1069
Q	y	2133	GAAGGAGGGTTGGAGGACCGAGGCACCATGGGGAACCGGCCCTCTTCCGGTGTCTC	2192
D	b	1070	-----GlyGlySerThrAlaPro-ProProProProProProProProProProPro	1085







```
Db 1109 GlyAlaArgGlyLeuAenGlyProGlnGlyProArgGlyAspGlyGlyAlaGlyGlu 1128
QY 1496 -----AGCGGTGAGTATAGG---TTTCTGGATGAGGGGCGAGCCCTG 1452
Db 1129 AlaGlyGluArgGlyGlnLysGlyHisArgGlyPheThrGlyLeuGlnGlyLeu--ProG 1148
QY 1451 GGTCCCTGTGCGATTGAGAGCTGGCACTGCAGGAGGTGCCAGAGACCCGTGCACAGC 1392
Db 1148 lyProPro-----GlyThrAlaGlyAspGlnGlyAlaThrGlySerAlaG 1163
QY 1391 ACCTAACATCCCCAC-----GAGTCCCTGTGGCAGGATGACAGCAAGCATGAGC 1338
Db 1163 lyProSerGlyProArgGlyProGlyProValGlyProSerGlyLysAspGlySerA 1183
QY 1337 GCAGC-----CGGGCCACACCGGGGAAGCA----- 1312
Db 1183 snGlylleProGlyProIleGlyProGlyProArgGlyArgSerGlyGluThrGlyP 1203
QY 1311 --GCTGGCAGCCCC-----TCAGGTGCTG-----TCCCCACCTTGGGCTGGT 1270
Db 1203 roAlaGlyProGlyGlnSerGlyProGlyProGlyProPro-----GlyP 1221
QY 1269 GGGGGTGTGGGTGGGACACAGATGGAGGACAGGGATGGAGGGGGAACACATCC 1210
Db 1221 roGlyIleAspMetSerAlaPheAlaGly-----L 1231
QY 1209 TGCAAGACGGGAGAAAGAGAGGCATGTGTATGTTTGGATGAGGACTTGGCAGCA 1150
Db 1231 euSerGlnProGluLysGlyProAspProMetArgTyrMetArgAlaAspGlnAlaSer 1251
QY 1149 TGAGTGCTATCAGTAGATCTTGTAGATGATCTGTGTGTGTCAGCTGAGAGAGTCTTTGG 1090
Db 1251 erSerValleuGlnThrAlaAspValGluAlaAsnLeuLysThrLeuAsnAsnGlnIleG 1271
QY 1089 CATGTGTG-----CCACTGTGTGTGTGCCAATG 1060
Db 1271 luSerIlelleArgSerProAspGlyThrLysLysAsnProAlaArgThrCysArgAspL 1291
QY 1059 TATATGTGTG-----ACATGAGGAGCGGGCCAGCGCTAGGAGCTGCGAGTCCGACT 1006
Db 1291 euLysLeuCysHisProGluTriLysSerGlyGluTyrTrpIleAspProAsnGluGlyC 1311
QY 1005 GC-----CTGTGTCCTGAGTTCACCTAGCTTCCTGCTGCGAGCTTT 967
Db 1311 yThrValAspAlaIleLysValPheCysAsnMetGluThrGlyGluSerCysValTyrP 1331
QY 966 CCAGTCCCACT 956
Db 1331 roSerProSer 1334

RESULT 3
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
R:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

Alignment Scores:
Pred. No.: 4,97e-13 Length: 1188
Score: 336.00 Matches: 278
Percent Similarity: 30.50% Conservative: 88
Best Local Similarity: 23.17% Mismatches: 321
Query Match: 3.56% Indels: 516
```

```
DB: 2 Gaps: 50
US-09-931-704-3 (1-5087) x S49915 (1-1188)
QY 15 CTTGCGGATGGGATTTAAAGCTTCGCGGAGCGCGCTCGCCCTCCACTCCGCCCA 74
Db 412 ProGlyGly-----GlyProProSerSerPro 420
QY 75 GCCTCCGGGAGGAGCGCACCCGGC---CGGCCAGCCCC---AGCCCCATGGACCTC 128
Db 421 ValProGlyLysProAlaAlaSerAlaProMetProSerProHisThrProAspVal 440
QY 129 CGAGCAGGTGAAAACCCAACTAGCCCTGCTCTTATAACATGACAAGCAGCGCCCAT 188
Db 441 SerProGluProLeuProGluProSerProVal-----ProAlaProAla 455
QY 189 CTGATACCTAAACCGACCAAGTCACAGC-----CCTCAACTCA 227
Db 456 ProMetProMetProThr-ProHisSerProProAlaAspAspTyrValProProThrPr 475
QY 228 CCTCTGCTGCCCGACAGCTCACCA----- 252
Db 475 oProValProGlyLysSerProAlaThrSerProSerProGlnValGlnProProAl 495
QY 253 -----CATCCTTTGTGACT 266
Db 495 aAlaSerThrProProSerLeuValLysLeuSerProProGlnAlaProValGlySe 515
QY 267 CAACCTCAACGGCACTAAATCAACAATCCCAAGTCTAAACTAATCTGAAACTTTTAA 326
Db 515 rProProProPro-VallYsThrThrSerPro----- 525
QY 327 AGTAACCCAGTCTTAAACCTAGCCCAATGCCAATTATATCTACCTAGCCAAAC 386
Db 526 -----ProAlaProIleGlySerProSerProProProValSerValSerPro- 543
QY 387 CTAACCTCCTTTGCCAGTCCAAAGTGTCCACTGAATCTCCTCTGGTCTCCTCACTGAA 446
Db 544 -----ProProProValLysSer-ProProProAlaProValGlySerProPro 560
QY 447 ATCCAGAAAGCATATTTCCCACTGCGCCATCCCTCTTACAGCAGCCACCTGGC 506
Db 561 ProProGluLys-----SerProProProAlaProValAlaSerProProProPro 578
QY 507 CTCTGGACT-----CCTGGTATCTGGGATGTCCAAACTCTGCAGTGCATCAGCC 557
Db 579 VallYsSerProProProThrLeuValAlaSerProProProProValLysSerPro 598
QY 558 AACAGCCCGACTCGTCAATGCACTCTCTCCCTCTGTCCTGTCCTGAGGCTGA 617
Db 599 ProProPro-----AlaProValAlaSerPro-ProProProValLys----- 612
QY 618 TGGAAAGGCCTCATTTGAAGTCCAACTTTTCCCACTTAACACACAGAAAGGGTGAACCT 677
Db 613 -----SerProPro-ProPro-----ThrProValAla 622
QY 678 CCACACTGCCACCGTTCCTGAGTGCAGCACTAAATCTCTTCAATCTAAACCCACCT 737
Db 622 erProProProProAlaProValAla-----SerSerProProPro 636
QY 738 ACATCTCCCACTCAGGAATCATCTAGTAATATACCAAACTAAGCCCATAGGC 797
Db 636 etLysSerPro-----ProProProThrProValSer 647
QY 798 AGCCGACCTAGTGGTCTAACCTATACCTTCTCTATGCTGGTGTCTGTCTTGGC 857
Db 647 erProProProProGluLys-----SerPro 656
QY 858 GCGCGCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 917
Db 656 roProPro-ProProAlaLysSerThrPro----- 665
QY 918 ATGTGCTGTCTCCCACTCTGACTCTCCCTCTCAAGCTGCAGTGGAGTGAAGACTGGCAG 977
```

```
Db 666 -----ProGluGluTy-Pro----- 671
QY 978 GAAGCTAGGTACAACTGGGAACACAGGAGGTGCGTCCAGTGCAGTCCCTAGGCTCGCCCGC 1037
Db 671 ----- 671
QY 1038 TCCCTCCATGTACACACATATACATGTGGCACACACACAGTGGCACACATGCCAAAGAC 1097
Db 671 ----- 671
QY 1098 TCTCTCAGCTGACACACAGATCCATCTCAAGTATCTACTGATAGACACTCATCGTGCC 1157
Db 672 -----ThrProThrSer----- 676
QY 1158 AAGTCCTCATCTCAACATACACATGCCTCTCTCTCTCCCGTCTTGCCAGGAGTGT 1217
Db 677 -----VallySerSer 680
QY 1218 TCCCTCTCCATCCCTCTCCCTCCATCTGGTGTCCACCCCTCAC-CCGCCACCCAGC 1276
Db 681 ProProGluLysSerLeuProPro-----ProThrLeuLeuProSerProPro 697
QY 1277 CCAAGTGGGACACACACTGAGGGGTGCGAGTGTCTCCCGTGGGGCC- 1330
Db 698 ProGlnGluLysProThrProProSerThrProSerLysProProSerSerProGluLys 717
QY 1331 -----GGGCGCGCTCATGTTCTCGTCCATCTCGCCACAGGGGACTCGTGGGGATG 1384
Db 718 ProSerProProLysGluProValSerSerProProGlnThrProLysSer----- 734
QY 1385 TTAGCGTCCGTGTGCAGCGTGTCTGGCACCTCCCTGCAGTGCAGTCTCAATCGCAC 1444
Db 735 -----SerProProProAlaPro-ValSerSerProPr 745
QY 1445 GGGACCCAGGCGCTGGCCCTCCATCCAGAAACCTACCTACCCGCTACCTGAG 1504
Db 745 oProThrProValSerSerProProAlaLeuAlaProValSerSerProProSer----- 763
QY 1505 CACCAACTCCGAGTGTGGTGGGACCTATGTGAGTATCCAGCTAGGAATCGGGAGTT 1564
Db 763 ----- 763
QY 1565 GGGGAGGAGTGAGGAGTTGGGGAAGACAGTCTTAACCGTGGAGGTTCTGGTAAATGAT 1624
Db 764 -----VallySerSerPro----- 768
QY 1625 GGGGTGAGGAGGCGCTTTTGGCTCCACAGTCCCGTGTGTCTATCTCTCGCCCT 1684
Db 769 -----ProProAlaProLeuSer-----SerPro----- 776
QY 1685 TCCCTCTTAGTGGGCGCCCGCTCCCATCCCTGGCCCGGAGTATGGGCA 1744
Db 777 -----ProProAlaPro----- 781
QY 1745 GGCCTCGCACCGCTTGGCCCATGGCCCACTGCTGCCAGCCAGCGCGCGCTCC 1804
Db 781 nVallySerSerProProProValGln-----ValSerSerPro-ProProAlaProL 799
QY 1805 CTGGGGCGGGGAAGTCTCTCTGTGTATACACCGTGTGTGTGTCTCTTGGCGGGCG 1864
Db 799 ysSerSerProProLeuAlaProValSerSerPro-ProGlnVal----- 813
QY 1865 GGGTGGTGGGACAGAGGGGCGCCCTCCATCGCTGCTCCAGTCCCTCTGCG 1924
Db 814 -----GluLysThrSerProProProAlaProLeu-SerSerProProLeuAl 829
QY 1925 CCGACAGCTGGGGCGCTGTCTGTGGACCCAGGGGCGCTCCCTCTCGCTGCTCTCCCA 1984
Db 829 aProLysSerSerPro-----ProHisValValValSerSe 841
QY 1985 TCCTAGCTGGGCGCTCTAGGGGGGTGTCATGGGGGAAGGGGACTGTAGGGAAACCCAGGCAGT 2044
|||
```

---

```
Db 841 rPro-----ProProVa 845
QY 2045 AGTGGCAGGGGTTTAGGGTGTGGATGAGGTATGCTGTAAAGATTTGGGGTGGTCCA 2104
Db 845 lVal----- 846
QY 2105 GAGGTGTTCAAGAGACCCAGGAGAGAGAGAGAGGGTGTGGAGGCGGAGCCACCATGG 2164
Db 847 -----LysSerSerPro----- 850
QY 2165 GGAACCGGCGCCCTCTTCCCGTGTCTTCCATCCACAGCCCTACTCTGGAGCCAG 2224
Db 851 -ProProAlaProValSerSerProProLeuThr---ProLysProAlaSerProProAl 869
QY 2225 GGAAGAAAAGGGAAGAGGTGGGGGAGCTGGCTCCAGCCCGCAGGATACACCGAGA 2284
Db 869 aHisValSerSerProProGluValValLysProSerThrProProAlaProThrThVa 889
QY 2285 AATTAGTTTGTCTGTGTGTGTGTCAGCGTGTGAACCTCCCTCCCTGGGCCCTT-GCCTATCC 2343
Db 889 lIleSer-----ProProSerGluProLysSerSerPr 900
QY 2344 CAGGCTCTCCCTTGTCTTCTCCCTTCTTCCAGTATTATATCTCCCTCATCTCCCTTCC 2403
Db 900 oProProThrProValSerLeuPro-----ProProLleVallySerSerProProPr 918
QY 2404 CTGGGCGCCAGCGCTCCCGAGGGTGTGAAGGGCTCTGCCCTCTTCCCTATACCATG 2463
Db 918 oAlaMetValSerSerProProMetThrProLysSerSer-Pro----- 932
QY 2464 CTGTCTTCCATAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2523
Db 933 -----ProProValValValSerSerProProProThrVallySerSers 947
QY 2524 CAACCTGTCTCTATCAGCTGAACCTCTTTCGGAGTGTAGTAGTACCGTCTCTCTCC 2583
Db 947 er-----ProProProAlaProValSerSerProProAlaThrP 960
QY 2584 CCAGCGCTCAGCTGTGGGCTGGGTGTGCAGCGGCAATGGGGCTCTGGTCCAATG 2643
Db 960 roLys---SerSerProProProAlaProVal-----AenL 971
QY 2644 GGCACCTCTCATCTCTCTCTTGTTCGAGAAACCTTGTCTCAGTCCACTGCC 2703
Db 971 euProProGluVallySerSer-----ProPro-----ProThrProV 985
QY 2704 TCTCTAGTTCGCGACCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2763
Db 985 alSerSerProProPro-----AlaProLysSerSers 995
QY 2764 GTCTACACCTCTGCT-----CCACTTCTCTC 2790
Db 995 erProProProAlaProMetSerSerProProProGluVallySerSerProProProp 1015
QY 2791 TCCACCCACTCACTTCTTAACCCCTGCAATCTGGCTTCCAGGCCCGCAAGTGTCTC 2850
Db 1015 roAlaProValSerSerProProProVallySerSerProProPro-----AlaP 1033
QY 2851 TCCAGGTGTCAGGCACTCTTGGCCAGGCCGACAGTGTGTGAAGGCTCATCTCTCT 2910
Db 1033 roValSerSerProProProVallySer----- 1043
QY 2911 TGCTGTCTGTTTGGCAGCCACACTGCTGAGCGCTGCTCTCTCTCTCTCTCTCTCTCTCT 2970
Db 1044 -----ProProProProPro 1048
QY 2971 GGTCTCTGCACTCTCTGGGCCACCTTACCTCTCCAGCTCTCTCAGGCTCTCTCTCTCT 3030
Db 1048 laProValSerSerProProProProVallySerSerProProProAlaProLleSerS 1068
QY 3031 CTCTCTCTGCGCCCGCACAGCGGGCACTCTCCCAAGTGTGGCCACCGCCAGCCAGCAGC 3090
Db 1068 er-----ProProProVallySerSerp 1076
|||
```











QY 3800 TCTTGAGCCGGTTGAAGTCTCTGGCCGAGCCACAGCCAGGTCTGACAGTCTCTTCACCA 3741  
Db 364 -----SerAlaGlyProSerGly 369  
QY 3740 GCACAGAGTCTGCATCTCTGGAGGAAGTCACTGTGGGAGGCCAGGAGTCCAAAGTGG 3681  
Db 370 AlaGlnGlyThrProGlyPheProGlyAlaAargGlyProProGlyProAlaGly----- 387  
QY 3680 GTTCAGTCCACGAGCAGCGGTGGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3621  
Db 388 -----SerProGlyProAlaGlySerLysGlyAsp 398  
QY 3620 TGCTGTCGACGAGCCCTGGAGGCTGGTGCAGAAAGTGGGCGAGGCTGGGCGCAGCTCAG 3561  
Db 399 GlyAsnProGlyGlnProGlyAlaGlnGlyGlySerGlyPro---LeuGlyProAargGly 417  
QY 3560 CAGTGGCAGCTGACGGTGGAGGCCACGCAAGTAAACAGAGTGGTGGTGGTGGTGGTGGTGG 3501  
Db 418 GlnThrGlyProAlaGly-----ProProGlyAlaGlnGlyGlySerGlyGlyGluAarg 434  
QY 3500 AGTTCTGGGTGACGCGAGTTGTGTCATTGAGGCTTCGCCACACCTCCCAAGTCAACAGTGG 3441  
Db 435 GlySerAargGlyAlaLeu-----GlyProAlaGlyProProGlyGlyValGly 450  
QY 3440 CCTGGGCGAGAGTCTCTGCCCGGAGGCGGAGGTTGAAGTCTGGCTCGTTGAAAGGGG 3381  
Db 451 GluAargGlyProMetGlyProProGlyMet-----SerGly 462  
QY 3380 GGCACAGGTAGTTCAGCTGTGAAA-----AGGAGGGGTGATGG 3342  
Db 463 AlaProGlyAlaProGlyAlaLysGlyAspAargGlyLeuProGlyGluAargGlySerAla 482  
QY 3341 GGAAGG-----AGGAGGCGAGAGCGCTGGCTCACCACCA 3306  
Db 483 GlySerLysGlySerAlaGlyGlySerGlyAargProGlyGluPro---GlyMetProGlyG 502  
QY 3305 GGAGATACCTGCTCCCAAGTTCTATTGTTGTCTCCCTGACACTGGCCCTGCTCCA 3246  
Db 502 LnaAargGlyLeu-----ThrGlyProProGlyL 511  
QY 3245 TGCTAGGCTCTTGTGTCATGCTGCTCCTGAGTTTTCATAAGGATATCATTTGTATG 3186  
Db 511 ysGlnGly-----AargAspGlyLysPro-----GlyProAlaGlyAlaP 524  
QY 3185 CAGGTGACAGTTCAGAGCAGTGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3127  
Db 524 roGlyGluProGlyAenSerGlyProAlaGlyAlaSerGlyGlnAargGlyLeuProGlyL 544  
QY 3126 GAGGAGGAGGAGGACCAAGAGCTCAGGAAGAGCTGTGATGCTGGTGGTGGGCAAA 3067  
Db 544 euValGlyLeuProGlyProGlnGlyGlnAargGlyGluAargGlyGluAargGlyGlnG 564  
QY 3066 CCTTGGGAGAGTCCCG---CTGGGGGCGAGCAGACAGAGGAGAGAGGAGGAGGAGGAG 3010  
Db 564 lysSerProGlyAlaProGlyLeuThrGlyGluProGlyLysAargGlyGlyProGlyVala 584  
QY 3009 CTGGAGAG-----GTAGAAAGTGGCCGAGGAGAGTGCAGACCAAGGAGAGGAGTTC 2956  
Db 584 laglyProProGlyProGlnGlySerAlaGlyGluAargGlyAenGlnGlyProGlnGlyG 604  
QY 2955 GAGAAAGCAGC-----ACGCTCAGCAGTGTGGTGGTGGCAAAA 2920  
Db 604 lnaAlaGlySerMetGlyProProGlyProProGlyAlaSerGlyAspAlaGlyAlaGlnG 624  
QY 2919 CAGACAGCAAGGAGGAATGAGCCTTCAAAACACTGTCGGGCTTGGCAGGAGGTCCCTGAC 2860  
Db 624 lysAsn----- 626  
QY 2859 GACCTTGGAGAGAACCATTTCTGGGCGCTGGAAGCCAGATTGACGGGGGTTAAGAGTGA 2800  
Db 627 -----GlyProProGlyGlu-----S 632

QY 2799 GTGGTGGAGAGCA---AGTGGAGGCGAGGAGGTGTAGACCACCTCTTGGAGAAAATTTGGC 2743  
Db 632 erGlyProGluGlyProProGlyAlaAargGlyGluAarg----- 644  
QY 2742 AGGAAAGCCAGGAGAGAAA-----AAGGTGCGGAACCTAGAGAGGCGAG 2698  
Db 645 --GlyAlaProGlyGluAargGlyProGlnGlyLeuThrGlyAlaGlnGlyAargGlyI 664  
QY 2697 TGGAGTGAAGCAAAAGTTTCTGCACAAGGAACAAGAGAGATGAGATGGCCCATGG 2638  
Db 664 leGlyGlyAargAlaGlyAsnTyrGlyAlaThrGlyProProGlyGlnLysGlyGluMetG 684  
QY 2637 AACACAGAGCCCATTTGGCGCTGCACACACCCAGGCGCCACAGCTGAGGCGCTGGGAGAG 2578  
Db 684 lyProProGlyAsnValGlyLeu-----GlnGlyProProGlyGlu---LeuGlyProS 701  
QY 2577 ACGGTACTCTACTAATCACTCCGAAAAGAGGTTTCAGT-----GATAGGA 2533  
Db 701 erGlyProProGlyAlaAargGlyProProGlyProSerGlySerProGlyProAspGlyP 721  
QY 2532 GCAGGTTTGCAGAAAGAAATGAGGCGAGTCTCATGATAGTAGGACAGGAGGAAGGCTAT 2473  
Db 721 roAlaGlyAlaGluGlyAspAargGly-ProValGlyProMetGlyProSerGlyProSer 740  
QY 2472 GGAAGACAGCATGTATAGGGAAGAGGCGAGAGCCCTTTCCAACTCCGGGGAGCG--- 2416  
Db 741 GlyMetProGlyGluAargGlyAspAenGlyGluProGlyPro-----GlyGlyAlaPro 758  
QY 2415 -----GCTGGGCGCCAGGAAAGGATGAGGGAGATGTATA 2380  
Db 759 GlyGlnAargGlyAspProGlyAspLeuGlyProGlnGlySerProGlySerProGlyPhe 778  
QY 2379 ACTGGAAAGAGGAGAGCAAGGAGGAGAGC-----CTGGATAGGCAAGG 2332  
Db 779 AlaGlyProProGlyAargSerGlyAsnProGlyProGlnGlyGluLeuGlyProThrGly 798  
QY 2331 GCCAGGGGAGGTTTCACACGCTGACAAGACAGACAGACAACTAATTTCTCGGTGTATC 2272  
Db 799 AlaAargGlyGluThrGly----- 804  
QY 2271 CTGGGCTGGAGCCAGCT-----CCCCCGCCACCTTCTTCCCTTTCTTT 2227  
Db 805 ---GlyProGlyProSerGlyProThrGlyAspProGlyProGlnGlyProLeuGlyAla 823  
QY 2226 CCTTGGCTCCAGAGTAGGCTCTGGATGGAAGAGGACACGGAAGAGGCGCGGTT 2167  
Db 824 ProGlyGlnGlnGly-----GluAargGlyGluThrGly----- 834  
QY 2166 CCCATGTGCTCGGCTCTCCAAACCTCTTCTTCTCTCTCTGCGGTCTCTGAACACC 2107  
Db 835 ProGlnGlyGlnGlyGlyPro-----ProGlyProIleGlySerLeuGlyAla 850  
QY 2106 TCTGGACCA-----CCCCAAATCTTACACATACCTCCATCCACACCCTAAACCCC 2053  
Db 851 ProGlyAlaGlnGlyProProGlyPro-ThrGly-----ProSer----- 863  
QY 2052 CTGCCACTACTCTCGTGGTTCCTCAGTCCCTTCCCTCCCATGACCCCCCTAGAGGCC 1993  
Db 864 -----GlyAsnAlaGlySerProGlyGlnProGlyAlaAargGlyGluPr 878  
QY 1992 AGCTAGTAGTGGAGAGGCGAGC-----GGAAGGAGGCCCT---GGTCC 1951  
Db 878 oGlyGlnSerGlySerProGlyGlnProGlyLeuAlaGlyAargThrGlyProSerGlyG 898  
QY 1950 CAGAGCAGAGGCGCCCGAGGCTGGGGCGAGGCGAGCTGGGAACGACGATGGGAGGT 1891  
Db 898 uAarg-----GlyAspLysGlyAsnAserGlyGlnSerGlyProPr 911  
QY 1890 GGGGCGCTCTGTCTCCCAACCCCAACCCCGCGCAAGAGACACACACACAGCTGTAAA 1831  
Db 911 oGlyPro-----ProGlyProAlaGlyProAlaGlyGlnSerGlyIleLeuGlyLeuAl 929  
QY 1830 CAGAGGAGACTTCCCC-----GGCCCCAGGGGAGGCGGGC-----GGCTGGGCTGGCAG 1780





Qy	1339	CGCGGCGCCG--	-GGCCACACAGGGGAGCAGCTGGCGAGCCCTCAGCTGTCTGTCTCCCA	1283
Db	1092	uSerGlyProGlnGlyProLeuGly	-----ProSerGlyProAlaGlyAl	1107
Qy	1282	CCTTGGGCTGGTGGG	-----GGTCAGGTGGCAGCACCAG	1247
Db	1107	aArgGlyLeu-AlaGlyProGlnGlyProArgGlyAspGlyGluAlaGlyGluAlaG	-----C	1211
Qy	1246	ATGGAGGCACAGGGGATGGAGGGGAAACACT		
Db	1127	lyGluArgGlyGlnGlyHisArgGlyPheThrGlyLeuGlnGlyLeuProGlyProP		1147
Qy	1210	CTGGCAACAGCGGAGAGAA	1191	
Db	1147	roGlySerAlaGlyAspGln	1153	
RESULT 7				
T45025				
mucin MUC5B, tracheobronchial [imported] - human (fragment)				
C:Species: Homo sapiens (man)				
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000				
C:Accession: T45025				
R:Dessevyn, J.L.; Guyonnet-Duprat, V.; Porchet, N.; Aubert, J.P.; Laine, A.				
J. Biol. Chem. 272, 3168-3178, 1997				
A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes vari				
A:Reference number: 222899; MUID:97166151; PMID:9013550				
A:Accession: T45025				
A>Status: preliminary; translated from GB/EMBL/DBJ				
A:Molecule type: DNA				
A:Residues: 1-3570 <DES>				
A:Cross-references: EMBL:Z72496; NID:gl834502; PIDN:CAA96577.1; PID:gl834503				
A:Experimental source: placenta				
A:Genetics:				
C:Gene: MUC5B				
Alignment Scores:				
Pred. No.:	1.35e-12	Length:	3570	
Score:	329.50	Matches:	432	
Percent Similarity:	30.37%	Conservative:	179	
Best Local Similarity:	21.47%	Mismatches:	734	
Query Match:	3.49%	Indels:	671	
DB:	2	Gaps:	90	
US-09-931-704-3 (1-5087) x T45025 (1-3570)				
Qy	195	CTAAACCGACCAAGTCACAGCCCTCCAACTCACCTCTGCTGCTGCCAGCTCACCACA	254	
Db	1619	ProSerThrProAlaThrSerSerThrAlaThrProSerSerThrProGlyThrThrTip	1638	
Qy	255	TCCTTTGTGACTCAACCTCAACCCGACTAAATCAACCAATCCCAAGTCTAACTAATC	314	
Db	1639	lLeuThrGluGlnThrThrAlaAlaThrThrAlaThrThrGlySerThrAlaIle	1658	
Qy	315	TGAAACTTTTAAAGTAACCCAGCTCTTAAACCTTAACCTAGCCCAATGCCAATTTATATCTA	374	
Db	1659	-----ProSerSerThrProGlyThrAla-----	1666	
Qy	375	CCCTAGCCAAACCTTAACCTGCCTT-----TCCAGTCC	407	
Db	1667	-----ProProProlysValLeuThrSerGlnAlaThrThrProThrAlaThrSerSer	1684	
Qy	408	AAAGTCTCCACTGAATCTCCTCAGCTCTGCTCCTCAGTGAATCCAGAAAAGCATATTTCC	467	
Db	1685	LyAlaThrSerSerSerSerProArgThrAlaThrThrLeu-----	1698	
Qy	468	CCACTGCCACATCCCTCCTTACAGCACCCCAACCCCTGGCTCTGGACTCTGGTATCTCTG	527	
Db	1699	ProValLeuThrSerThrAlaThrLysSerThrAlaThrSerPheThrPro-----	1715	
Qy	528	GGATGTCAAACTCTGCAGT-----CCATCAGCCCAACAGCCGACTCGTCA	575	
Db	1716	-----lLeProSerSerThrLeuGlyThrThrGlyThrSerGlnAsnArgProHisPro	1734	

QY 576 AATGCACTCTCTCCCTCTCTGTCCTCCACCCCTTGACGCTGATGGAAGGCTCATTTGAA 635  
Db 1735 MetAlaThrMetSerThrIleHisProSer----- 1744  
QY 636 GTCCAACTTTTCCCACTTAACACCAAGAGGGGTGACCTCCACACTGCGACCGTTCC 695  
Db 1745 -----SerThrProGluThrThrHisThrSer 1753  
QY 696 CTGAGAGTGAGCACTAA-----TCTCCTTCA 722  
Db 1754 ThrValLeuThrThrLysAlaThrThrArgAlaThrSerSerMetSerThrProSer 1773  
QY 723 ATCTAACCCACCCCTACACTTCCCACTCCAGCAATCATCTCAGCAATATACCCAAAC 782  
Db 1774 SerThrProGlyThrThrTrpIleLeuThrGluLeuThrThrAlaAlaThrThrAla 1793  
QY 783 TAAGCCCAATAGGACGCGCCGACC-----CTAGTGGTCTAACCC 821  
Db 1794 AlaLeuProHisGlyThrProSerSerThrProGlyThrThrTrpIleLeuThrGluPro 1813  
QY 822 TATACCTTGCTTCTATGGGTGAGTCTGTTCTTGGCGCGCCCTCTCTCTCTCTCTCTCC 881  
Db 1814 SerThrAlaThrVal-----ThrValProThrGlySerThrAlaThrAlaSerSer 1831  
QY 882 CTTAGAGCTGAGTGTCTAGCTGCGCAGCTCTGACAGTGTCTGCTCTCCACCTT----- 936  
Db 1832 ThrArgAla-ThrAlaGlyThrLeuLysValLeuThrSerThrAlaThrThrProThrVa 1851  
QY 937 -----CTGACTCCCTCAAGCTGACGTGAGCTGGAAGA-----CTGG 974  
Db 1851 lIleSerSerArgAlaThr-ProSerSerSerProGlyThrAlaThrAlaLeuProAlaL 1871  
QY 975 CAGGAAGCTAGGTACAACTGGAACACAGGAGGTGCGACCTGCACTGCTCCCTAG----- 1026  
Db 1871 euArgSerThrAlaThrThrProThrAlaThrSerValThrAlaIleProSerSerL 1891  
QY 1027 -----GCTGGCCCCCTCCCTCCATGTACACA-----CATATACATGTGGACAC 1072  
Db 1891 euGlyThrAlaThrThrArgLeuSerGlnThrThrProThrAlaThrMetSerThr 1911  
QY 1073 ACACAGTGGCACACATGCCAAAGACTCTCTCAGCTGACACACACATCCATCTCAAGTAT 1132  
Db 1911 laThrProSerSerThrProGluThrValHisThrSerThrValLeuThrThrThrAla 1931  
QY 1133 CTACTGATGACACTCATGCGTG-----CAAAGTCTCATCTCT----- 1170  
Db 1931 hrThrThrArgThrGlySerValAlaThrProSerSerThrProGlyThrAlaHisThr 1951  
QY 1171 -----CAAACATACATGCTCTCTTCTCTCCGCTTCCAGGAGTGTTCCTCC 1222  
Db 1951 hrLysValProThrThrThrThrThrGlyPheThr-----AlaThrPro--SerSerSerPr 1969  
QY 1223 TCTCCATCCCTCTGCTCCCTCTGCTGTCCTCCACCTCACCCCTCCACCCAGCCCAAGG 1282  
Db 1969 oGlyThrAlaLeuThrProProValTrpIleSerThrThrThrThrThrThrThr 1987  
QY 1283 TGGGGACAGACACTGAGGGGCTGCCAGCTGCTTCCCGTGTGGGCGGCGCGCTCA 1342  
Db 1988 -----ArgGlySerThrValThrProSerSerIleProGlyThrThrHi 2002  
QY 1343 TGCT----- 1354  
Db 2002 sThrAlaThrValLeuThrThrThrThrThrThrValAlaThrGlySerMetAlaThrPr 2022  
QY 1355 ATCTCTCCACAGGGGACTCGTGGGGATGTAGCGGTGCTGTCACGGTGTCTGCGAC 1414  
Db 2022 oSerSerSerThrGlnThrSerGly-----ThrProPr 2033  
QY 1415 ATCTCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGGCTGGCCCTCCATCCAG 1474  
Db 2033 oSerLeuThrThrAlaThrThrIleThrAlaThrGlySerThrThrAsnProSerSe 2053  
QY 1475 AAAACCTATGACCTCACCC-----GCTACCTGGGAGCACCACTCCGAGC 1519

Db 2053 rThrProGlyThrThrProIleProProValLeuThrThrAlaThrThrThrProAlaAl 2073  
QY 1520 TTGGCTGGGACCTATGTAGTATCCAGCCTAGGAATCTGGGAGTTGGGGAGGTAGGA 1579  
Db 2073 aThrSerSerThrValThrProSerSerAla-----LeuGlyThrHis-- 2088  
QY 1580 GTTGGGGAAGACAGTCTCTAA-----CCGTGGAGGGTCTGTAAATGATGGGTGAGGAG 1635  
Db 2089 -----ThrProProValProAsnThrThrAlaThrThrHisGly-- 2101  
QY 1636 GGGCTCTTTGGCTCCCAACAGT-----CCCTCTGTCTGTCTATCTCTGCCCC 1683  
Db 2102 -ArgSerLeuProProSerSerProHisThrValProThrAlaThrThrSerAlaThrSe 2121  
QY 1684 TTCCCTCTTAGT-----GGCCCCCCTCTCCCATCTCCCTGCCCCCAGG 1728  
Db 2121 rGlyIleLeuGlyThrThrHisIleThrGluProSerThrGlyThrSerHisThrProAl 2141  
QY 1729 A-----CTAGCATGTGGCAGGCTCGCACCCGCTTGGCC-----CATTTGCCCCAC 1776  
Db 2141 aAlaThrThrGlyThrThrGlnProSerThrProAlaLeuSerSerProHis--ProSe 2160  
QY 1777 TGGCTGCCAGCCAGCCGCTCCCTCCCTCGG-----GGCCGGGG 1818  
Db 2160 rSerArgThrThrGluSerProProSerProGlyThrThrThrProGlyHisThrArgL 2180  
QY 1819 AAGTCTCTCTGTTTACACCGTGTGTGTCTCTTGGCGCGCGGG-----TTGGG 1872  
Db 2180 yThrSerArgThrThrAlaThrAlaThrProSerLysThrArgThrSerThrLeuLeuPr 2200  
QY 1873 TGGGACAGAGGGGCCCCACCTCCCATCGCTCGCTTCCAGCTCGCTCGCCCCCAGACC 1932  
Db 2200 oSerSerProThrSerAlaProIleThrThrValValThrThrGlyCysGluProGln-- 2219  
QY 1933 TGGGGCCTCTGCTCTGACCCAGGGGCTCTCTTCCCTCTGCTCTCTCCATCTAGCT 1992  
Db 2220 -----CysAlaThrPserGluThrLeuAspThrSerThrProMetPro----- 2233  
QY 1993 GGGCTCTCTAGGGGTCATGGGGAGGGACTGTAGGGAACCCAGGAGTAGTGAGCAG 2052  
Db 2234 -GlyProSerGlyGlyAspPhe-----AspThrTyrSerAsnIleArgAlaAlaGly-G 2251  
QY 2053 GG-----GCTTAGGCTGT-----GGATGGA 2073  
Db 2251 lAlaValCysGluGlnProLeuGlyLeuGluCysArgAlaGlnAlaGlnProGlyValP 2271  
QY 2074 GGTTA-----TGCTGTAAAGATTTGGGGTGGTCCAGAGGT 2109  
Db 2271 roLeuArgGluLeuGlyGlnValValGluCysSerLeuAspPheGlyLeuValCysArgA 2291  
QY 2110 GTTCAGAGAGCCAGGAGAGAAGGAGG-----AGGTTGGAGGAG 2150  
Db 2291 snArgGluGlnValGlyLysPheLysMetCysPheAsnTyrGluIleArgValPheCysC 2311  
QY 2151 CCGAGGACCATGGGGAACCGGCCCCCTCTTCCCGTGTCTCTCCATCTCCAG----- 2206  
Db 2311 ysAsn-----TyrGlyHisCysProSerThrProAlaThrSerSerThrAlaThrPro 2329  
QY 2207 -----ACCTACTCTGGAGCCGAGGAGGAGGAGG 2237  
Db 2329 erSerThrProGlyThrThrTrpIleLeuThrLysLeuThrThrThrAlaThrThrThrG 2349  
QY 2238 AAGAAGGTGGGGGAGCTGGCTCCAGC-----CCAGGATACCCGAG 2282  
Db 2349 luSerThrGlySerThrAla-ThrProSerSerThrGlnGlyProProAlaGlyThrPro 2368  
QY 2283 GAAATAGTTTG-----TCTGTGCTTGTTCAGCGTGTGAACCTCCCCCT 2327  
Db 2369 HisValSerThrThrAlaThrThrProThrValThrSerSerLysAlaThrProPheSer 2388  
QY 2328 GGGGCC-----TTGCCTATCCAGGCTCTCCCTTGCTTCTCCCTCTT 2372



Db 2389 SerProGlyThrAlaThrAlaLeuPro-AlaLeuArgSerThrAlaThrThrProThrAl 2408  
Qy 2373 TCCAGTATACATCTCCCTCATCCCTTTCCTGGGCCCCAGCGCTCCCGAGGGTTG 2432  
Db 2408 aThrSerPheThrAlaIleProSerSerSerLeuGlyThr-----Tr 2423  
Qy 2433 GAAAGGGCTCTCCCTCTCCATATACCATCTGCTCTCCATAGCCTTCTCTCTCTCTA 2492  
Db 2423 pThrArgLeuSerGlnThrThrThrProMetAlaThrMetSerThrAlaThrProSerSe 2443  
Qy 2493 CTCATGAGACTGCTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2552  
Db 2443 rThrProGlyThrValHis----- 2449  
Qy 2553 TTTCGGAGTGTAGTAGTACCGCTCTCTCCAGCCCTCCAGCTGGGGCGCTGGGT 2612  
Db 2450 -----ThrSerThrValLeuThrThrThrAlaThrThrGlyAlaThrG1 2465  
Qy 2613 GTCAGCGGCAATGGGGCTCTGGTTCCCAATGGGCACCTCTCATCTCTCTCTTTCTCTG 2672  
Db 2465 ySerValAlaThrProSerSerThrPro--GlyThrAlaHisThrThrLysValProTh 2484  
Qy 2673 TGCAGAAACCTTGCTCTCACTCCACTGCTCT-----CTAGTTCC 2714  
Db 2484 rThrThrThrThrGlyPheThrValThrProSerSerSerProGlyThrAlaArgThr-p 2504  
Qy 2715 CGACCTTTTCT 2774  
Db 2504 roProValTriPileSerThrThrThrThrProThrThrSerGlySerThrValThrProS 2524  
Qy 2775 CTGCC-----TCCACTTCTCTCCACCCACTCTCTCTTAAC-----CCCTGCA 2819  
Db 2524 erSeriLeProGlyThrThrHisThrProThrValLeuThrThrThrThrGlnProValA 2544  
Qy 2820 ATCTGGCTTCCAGGCCCGAGCAATGGTTCTCTCCAGGTGCTGAGGACCTCTCTGCCAA 2879  
Db 2544 laThrGlySerMetAlaThrProSerSerSerThrGlnThrSerGlyThr-----P 2561  
Qy 2880 GCCGACAGTGTTTGAAGGCTCATCTCTCTGCTGCTGTTTTCAGCACACTGCTGA 2939  
Db 2561 roProSerLeuThrThrAlaThrThrIleThrAlaThrGlySerThrThrAsnProS 2581  
Qy 2940 GCGCTGCTGCTCTCGAACT 2978  
Db 2581 erSerThr-ProGlyThrThrProIleProProGluLeuThrThrThrAlaThrThrPro 2600  
Qy 2979 -----CACTCTCTCTGG 2990  
Db 2601 AlaAlaThrSerSerThrValThrProSerSerAlaLeuGlyThrThrHisThrProPro 2620  
Qy 2991 CCACCTCTTACCTCTCCAGCTCTCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3047  
Db 2621 ValProAsnThrThrAlaThrThrHisGlyArgSerLeuSerProSerProHisThr 2640  
Qy 3047 ----- 3047  
Db 2641 ValArgThrAlaTriPThrSerAlaThrSerGlyThrLeuGlyThrThrHisIleThrGlu 2660  
Qy 3048 -----AGCGGGCACTCTCCCAAGTT-----TGC 3071  
Db 2661 ProSerThrGlyThrSerHisThrProAlaAlaThrThrGlyThrThrThrThrThrThr 2680  
Qy 3072 CCACCCAGCAATCAGACGCTCTCTCTGAGCGTCTGTGGCTCTCTCTCTCTCTCTCTCT 3128  
Db 2681 ProAlaLeuSerSerProHisProSerSerArgThrThrGluSerProProSerProGly 2700  
Qy 3129 TTTTCTACGCTCTCCATTTGAGAGCTCACCCGCACTGCTTCAACTGT----- 3179  
Db 2701 ThrThrThrProGlyHisThrThrAlaThrSerArgThrThrAlaThrAlaThrProSer 2720  
Qy 3180 CACCTGCATACAAATGATATCTTATTGAAAA----- 3212  
Db 2721 LysThrArgThrSerThrLeuLeuProSerGlnProThrSerAlaProIleThrThrVal 2740

Qy 3213 ---ACTCAGGAGGAGCCATGAACAAAGAGCCTAGCATGGACAGGCG----- 3257  
Db 2741 ValThrThrGlyCysGluProGln-----CysAlaTriPThrLeuAspThrSer 2758  
Qy 3257 ----- 3257  
Db 2759 TyrProMetProGlyProSerGlyAspPheAspThrThrThrSerAsnIleArgAlaAla 2778  
Qy 3258 -----CAGTGTCTCAGGAGCACAACAAA----- 3278  
Db 2779 GlyGlyAlaValCysGluGlnProLeuGlyLeuGluCysArgAlaThrAlaGlnProGly 2798  
Qy 3279 -----ATAGAACTTTGGGAGCAGGTATC-----TCCTTGTGTGTG 3314  
Db 2799 ValProLeuGlyGluLeuGlyGlnValValGluCysSerLeuAspPheGlyLeuValCys 2818  
Qy 3315 AGCCAGCGG----- 3323  
Db 2819 ArgAsnArgGluGlnValValGlyLysPheLysMetCysPheAsnTyrGluIleArgValPhe 2838  
Qy 3324 -----CTCTGCCCTCCT-----CCTTCC 3341  
Db 2839 CysCysAsnTyrGlyHisCysProSerThrProAlaThrSerSerThrAlaMetProSer 2858  
Qy 3342 CCATCACCT 3377  
Db 2859 SerThrProGly-----ThrThrTriPileLeuThrGluLeuThrThrThr 2873  
Qy 3378 -----GCCCTCTTCAACGAGCAGACTTCA 3404  
Db 2874 AlaThrThrThrAlaSerThrGlySerThrAlaThrProSerSerThrProGlyThrAla 2893  
Qy 3405 ACCCTCCCGCTGGGGCAGAGACTCTGCCAGGCGCACTGTTCTACTTGGAGGTGGC 3464  
Db 2894 ProProPro-----LysValLeuThrSerProAlaThrThrProThrAlaThr 2909  
Qy 3465 GAAGCCTCAATGACAACTGGCGGTGACCCAGAACT----- 3500  
Db 2910 SerSerLysAlaThrSerSerSerSerProArgThrAlaThrLeuProValLeuThr 2929  
Qy 3501 -----ACAGCGCTACAGCCCTCTGTTACTTGGTGGCGCTCAACCGCTC--- 3548  
Db 2930 SerThrAlaThrLysSerThrAlaThrSerValThrProIleProSerSerThrLeuGly 2949  
Qy 3549 -----AGGCTGCCACTGCTGAGCTGGCGCGCAGCCTGG 3581  
Db 2950 ThrThrGlyThrLeuProGluGlnThrThrThrProValAlaThrMetSerThrIleHis 2969  
Qy 3582 CCCACTTCTGACCAACCCCTCCAGGCG-----CTGCTGGGAGCATTCGGCGCTC 3631  
Db 2970 ProSerSerThrProGluThrThrHisThrSerThrValLeuThrThrLysAlaThrThr 2989  
Qy 3632 ATGGCAGCTCTGGGTACCCACTGCCAGCGCTGCTGGGACT----- 3676  
Db 2990 ArgAlaThrSerSerThrSerThrProSerSerThrProGlyThrThrTriPileLeuThr 3009  
Qy 3677 -----GAACCCACTTGG 3688  
Db 3010 GluLeuThrThrAlaAlaThrThrThrAlaGlyThrGlyProThrAlaThrProSerSer 3029  
Qy 3689 ACTCTGGCGCTGCCACAGTACTTCTCCAGAGATGAGCAGCTTCTGGCTGTGAAG 3748  
Db 3030 ThrProGlyThrThr-----TriPileLeuThr 3038  
Qy 3749 GAGCTGACAGCTGCTGGCGCTGGCCAGGACTTCAACCGGCTCAAGAGAGATG 3808  
Db 3039 GluLeuThrThr-Thr-----AlaThrThrThrAlaSerThrGlySerThrAlaThrLe 3056  
Qy 3809 CAGCTCCAGCAGCTGACCTGACCTGACCTGGGGGCTCATGGCTTCTGACT----- 3861  
Db 3056 user-SerThr-----ProGlyThrThrTriPileLeuThrGluProS 3070







QY 3101 TCAGGAAGGACGTGCTGATGCTGGTGGGCAAACTTGGGAGAGTCCCGCTGTGGG 3042  
Db 665 ----- 665  
QY 3041 GCAGGACAGAGGAGGAGGAGCTGGAGGAGCTGGAGAGGTAGAGGTGGCCCAAGGAGA 2982  
Db 666 -----GlyAlaProGlyAlaProGlyGlyLysGlyAspAla---GlyAlaProGlyGly 682  
QY 2981 GTGCAGAGACCAGGAAGAGGAGGTTGCGAGAAGGCGAGCGCTCAGCAGTGTGGCTGCAA 2922  
Db 682 uArgGlyProProGlyLeuAlaGlyAlaProGly----- 693  
QY 2921 AACAGACAGGAGGAGGAGGAGCTTCAAAACACTGTGGGCTTGGCAAGGAGGTGCTG 2862  
Db 694 -----LeuArgGlyGlyAlaGly 699  
QY 2861 ACGACTTGGAGAGAACCATTTGCTGGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2802  
Db 699 yProPro-----GlyPro-----GluGlyGlyLysGlyAl 709  
QY 2801 GAGTGGGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2742  
Db 709 aAlaGlyProProGlyProProGlyAlaAlaGlyThrProGlyLeu-----Gly 725  
QY 2741 GGGAAAGCCAGAGAGAGAA-----AAGGGTGGGAACTAGA 2706  
Db 725 nGlyMetProGlyGlyAlaArgGlyGlyLeuGlySerProGlyProLysGlyAspLysGlyGly 745  
QY 2705 GAGGCACTGAGTGAAGAAAGGTTTTCGCACAGGAGGAGGAGGAGGAGGAGGAGGAGG 2646  
Db 745 uProGlyGlyProGlyAlaAspGlyValProGlyLysAspGlyProArgGlyProThrGly 765  
QY 2645 CCCATTGGAACAGAGCCCCATTTCCGCTGACACACAGGCCCCCAGCAGTGGGGGCT 2586  
Db 765 yProLysGlyProProGly-----ProAlaGlyGlnProGlyAspLysGlyGlyGly-- 782  
QY 2585 GGGAGAGACGGGTACTCACTAACACTCCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2526  
Db 783 -GlyAlaProGlyLeuProGlyLeuAlaGlyProArgGlySerProGly--GlyArgGly 801  
QY 2525 TGCAGGAAGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2466  
Db 801 yGluThrGlyProProGlyProAlaGlyPheProGlyAlaProGlyGln-AsnGlyGly 821  
QY 2465 AGCATGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2406  
Db 821 roGlyGlyLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 833  
QY 2405 AGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2346  
Db 833 yGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 848  
QY 2345 TGGGATAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2286  
Db 848 erGlyProAlaGlyProProGly----- 855  
QY 2285 TTCCTCGGTATCTGGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2226  
Db 856 --ProGlnGlyValLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 875  
QY 2225 CTGGCTCCAGAGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2169  
Db 875 ro-----GlyAlaArgGlyLeuProGly 882  
QY 2168 TTCCTCGGTATCTGGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2109  
Db 882 lyProProGlySerAsnGlyAsnProGly--ProProGlyProSerGlySerProGlyL 901  
QY 2108 CCTCTGGACCCCAATCTTACAGCATTAACCTCCATCCACACCTTAACCCCTG 2049  
Db 901 yAspGly---ProProGlyProAlaGlyAsnThrGly----- 912  
QY 2048 CACTACTGCTGGGTTCCCTACAGTCCCTTCCCTCCCTGAGGAGGAGGAGGAGGAGGAG 1989

Db 913 -----AlaProGlySerProGlyVal-SerGlyProLysGlyAspAlaGlyGlnProGly 930  
QY 1988 AGGATGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1935  
Db 931 GlutylGlySerProGlyAlaGlnGlyProGly-----AlaProGlyProLeuGly 948  
QY 1934 -----CAGGTCTGGGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1881  
Db 949 IleAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProGlyProGlyMetProGlyPro 968  
QY 1880 ---TGTCCCAACCAACCCCGCGCAAGAGACACACCAACACGCGTGTAAAC----- 1830  
Db 969 GlySerProGlyProGlnGlyValLysGlyGlySerGlyLysProGlyAlaAsnGlyLeu 988  
QY 1829 -----ACAGAGACTTCCCGGCGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1785  
Db 989 SerGlyGluArgGlyPro---ProGlyProGlnGlyLeuProGlyLeuAlaGlyThrAla 1007  
QY 1784 GGCAGCCAGTGGGGCAATGGGCCAAGCGGGTGGCAGGCGCTGCCACACATGCCCTAGTCTCTG 1725  
Db 1008 GlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeuPro----- 1022  
QY 1724 GGGCCAGGAGTGGGAAGTGGGGGCGCCCTTAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1665  
Db 1023 -----GlyArg---AspGlySerProGlyGlyLysGlyAspArg----- 1034  
QY 1664 CAGGGGAGTGGTGGGAGCCCAAGAGCCCTCTCACCCCATCATTTACCAGAACCTCTCC 1605  
Db 1035 ---GlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisProGlyProGly 1053  
QY 1604 ACGGTAGGACTGTCTTTCCCACTCTCTCTCCCACTCTCTCTCCCACTCTCTCTCTCTCT 1545  
Db 1054 -----ProValGlyProAlaGlyLysSer 1061  
QY 1544 GGTACTCATACTAGTCCCAAGTCCGAGTGGTGTCTCCAGGTAGCGGGTGGAGT 1485  
Db 1062 GlyAspArgGlyGlySerProGlyProAlaGlyProAlaGlyAlaProGlyPro----- 1078  
QY 1484 CATAGTTTCTGGATGAGGAGG---CCAGGCGCTGGTCCCTGTCCGATGAGAGCT 1429  
Db 1079 -----AlaGlySerArgGlyAlaProGlyProGlnGlyProArgGlyAspLysGly 1095  
QY 1428 GGCAGTGGAGGAGTGGCAGAGCACCGTGCACAGGCGACGCTAACATCCCC----- 1378  
Db 1096 GluThrGlyGluArgGlyAlaAlaGlyIleLysGlyHisArgGlyPheProGlyAsnPro 1115  
QY 1377 -----CACAGTCCCTCTGGGAGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1330  
Db 1116 GlyAlaProGlySerProGlyProAlaGlyGlnGlnGlyAlaIleGlySerProGlyPro 1135  
QY 1329 ---GGCCACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1273  
Db 1136 AlaGlyProArgGlyProVal-----GlyProSerGlyProProGlyLysAspGlyThr 1153  
QY 1272 GGTGGGGGTGAGGAGT-----GGGACACAGTGGAGGAGGAGGAGGAGGAGGAGGAG 1219  
Db 1154 Ser-GlyHisProGlyProLleGlyProProGlyProArgGlyAsnArgGlyGluArgGly 1173  
QY 1218 AAAC-----ACTCTGGCAAGACGGGAGAG 1194  
Db 1173 ySerGluGlySerProGlyHisProGlyGln 1183

## RESULT 9

S59856  
collagen alpha 1(III) chain precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text\_change 13-Aug-1999  
C:Accession: S59856; S62120; S16373  
R:Roman, P.D.; de Crombrughe, B.  
Gene 147, 161-168, 1994  
A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete  
A:Reference number: S59856; MUID:95011609; PMID:7926795

A:Accession: S59856  
A:Molecule type: DNA  
A:Residues: 1-1464 <TOM>  
A:Cross-references: EMBL:X52046  
R:Toman, D.  
submitted to the EMBL Data Library, November 1994  
A:Reference number: S62120  
A:Accession: S62120  
A:Molecule type: DNA  
A:Residues: 1-866, 'G', 868-1464 <TOA>  
A:Cross-references: EMBL:X52046; NID:G575321; PIDN:CAA36279.1; PID:G575322  
R:Metaaranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
Biochim. Biophys. Acta 1089, 241-243, 1991  
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.  
A:Reference number: S16176; MUID:91274355; PMID:2054384  
A:Accession: S16373  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1442-1464 <MET>  
A:Cross-references: EMBL:X57983; NID:G50476; PIDN:CAA41048.1; PID:G50477  
C:Genetics:  
A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 211/3; 229/3; 247/3; 265/3; 283/3; 295/3; 673/3; 706/3; 742/3; 760/3; 776/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
C:Keywords: coiled coil; extracellular matrix  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-154/Domain: propeptide #status predicted <PRO>  
F:32-92/Domain: von Willebrand factor type C repeat homology <VWC>  
F:155-1464/Product: collagen alpha 1(I) chain #status predicted <MAT>  
F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Alignment Scores:  
Pred. No.: 2,42e-12 Length: 1464  
Score: 325.00 Matches: 377  
Percent Similarity: 28.83% Conservative: 73  
Best Local Similarity: 24.15% Mismatches: 549  
Query Match: 3.48% Indels: 562  
DB: 2 Gaps: 83

US-09-931-704-3 (1-5087) x S59856 (1-1464)

QY 4672 CCTACACCCCAAAATCTACTCATCTGTAACCTTTTCAGGAAATCTCTGTGTGGCTAT 4613  
DB 92 ProSerThrPro-----AlaProValLeuProAspGlyHis 103  
QY 4612 GGTGGCCCTCCCGAG-----CCTGGCAACCCACAGATACCTCTGGGAAAGG 4568  
DB 104 Gly-----ProGlnGlyProLysGlyAspProGlyProGlyIleProGlyArgAsn 121  
QY 4567 GGGCAGAGAGGCGACCGTAATCTGGAAGACAAATTCGAGGCAAGTCTCTGATGCTCAG 4508  
DB 122 GlyAsp-----ProGlyLeuProGlyGln 129  
QY 4507 CTCGGGTTTGTGTCACCTCTGTCTTGGAGACTCTGGTCTGTCTGTCTGTCTGTCTGT 4451  
DB 130 ProGlyLeuProGlyPro-----ProGlySerProGlyIleCysGlu 143  
QY 4450 CCTGCCCTGTGCTTTTCAGTATTTCTCTCTGTTTATAGGGCACTCAATGCGAGAC 4391  
DB 144 SerCysProThr-----PheAspSer-----GlyGlyGlnAsnTyrSer 153  
QY 4390 CCCAACCTTGCACTGTAATCTCATCTCTTTTGTGTAGAAATGCCACCATGTTTGTTC 4331  
DB 154 ProGln-----PheAspSer-----Val 165  
QY 4330 CTGAATGGATGATAGAGTATTTGGGGATAAGGGAGCAACCCCATCCAGGAGATG 4271  
DB 159 -----TyrAspValLysSerGly-----Val 165  
QY 4270 GGGCAGCCAAAGCGCTTACCCAGAAAGCATGTGTGTGTGGTGGCCCTCAACCCCTGC 4211  
DB 166 GlyGlyMetGlyGlyTyrPro-----GlyProAlaGlyProPro 178

QY 4210 CCTCCAGATGTGCCACCTGAGGGGCTGGTGGCCACTGGCCAAAGTGGTAGGCAAGCA 4151  
DB 179 GlyProProGlyProPro-----GlySerSerGlyHis-----ProGlySerProGlySer 195  
QY 4150 GGAGGCCATGAGCGGCTCTGACCAACCTGAACCACTTCACTCCTCGAGCATGACTT 4091  
DB 196 ProGlyTyrGlnGly-----Pro 201  
QY 4090 CCTGTAGAAACAGGACAGACAGAGCGGCTACTCTCTCTCTCTCTCTCTCTCTCTCTCT 4031  
DB 202 ProGlyGluProGlyGln-AlaGlyProAlaGlyProGlyProGlyProGlyProGly 220  
QY 4030 GGGAGGTGGGAGGAGACAGAGGCTCATCGCATCACACGCCCGGTCAGGAAAGGC 3971  
DB 220 uGlyProAlaGlyProAlaGlyLysAspGlyLysAspGlyLysAspGlyLysAspGly 240  
QY 3970 CAGAGGCTCACAGCTTCT 3923  
DB 240 uArgGly-----LeuProGlyProProGlyIleLysGlyProAlaGlyMetPr 256  
QY 3922 GGGCTGGCTCTCACAAAGTGGGAGCAGGGTTTGAAGGGGGGCGAAGAGAGAGAGTTCAG 3863  
DB 256 oGlyPheProGlyMetLysGlyHisArgGlyPheAspGlyArgAsnGlyGluLys-GlyG 276  
QY 3862 AAGTCAGAAAGCATGAGCCCGCAGGTGCGAGGTGACTGCGAGTGTGGAGGTGCATCTT 3803  
DB 276 lThrGlyAla-----ProGlyLeu-----LysGlyGluA 286  
QY 3802 CTTCTTGAGCCGGTGAAGTCTTGGCGAGCGCCACAGCCAGGTCTGCGAGCTCTTCAG 3743  
DB 286 sGlyLeuProGlyAspAsnGlyAlaProGlyProMetGlyProArgGlyAlaPro--G 305  
QY 3742 CAGCCAGAAAGTCGTCCATCTTCTGAGGAAGTCACTGTGGCGAGCGCCAGGATCCAAAGT 3683  
DB 305 lGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArgGlyAsnAsp----- 322  
QY 3682 GGGTTCACTCCAGACACCGCTGGGCGAGTGGGTAGCCAGAGTGCATGACGCCCGC 3623  
DB 323 -----GlyAlaArgGlySerAspGlyGlnProGlyProProGlyProProG 338  
QY 3622 AATGCTGC-----CCAGCAGGCCCTGGAGGCTGGTGCAGAAAGTGGGCA----- 3579  
DB 338 lYThrAlaGlyPheProGlySerProGlyAlaLysGlyGluValGlyProAlaGlySerP 358  
QY 3578 -----GGCTGGCGGCAGCTCAGCAGTGGCAGCTCAGCGTTTGGAGCCAGCAAGTA 3527  
DB 358 roGlySerAsnGlySerProGlyGlnArgGlyGluProGlyProGlnGlyHisAlaGlyA 378  
QY 3526 ACAGAAAGTGGTGTAGGCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 3489  
DB 378 lGlnGlyProGlyProGlyAsnAsnGlySerProGlyGlyLysGlyGluMetG 398  
QY 3488 -----GCCGCAAGTTTGTCTATTTGAGGCTTCCGCCACACCTCCAAGTCAAC 3446  
DB 398 lYProAlaGlyIleProGlyAlaProGlyLeuIleGlyAlaArgGlyProProGly- 416  
QY 3445 AGTGGCCCTGGCGAGTCTCTGCCCGAGCGGGAGGGTTCAGTCTGGCTCGTTGAA 3386  
DB 417 -----ProAlaGlyThrAsnGlyIleProGlyThrArgGlyProSer----- 430  
QY 3385 AGGGGGGCCAGGTAGTTTCACTGTGAAAGAGAG-----GGTGTAT-----GGG 3341  
DB 431 --GlyGluProGlyLysAsnGlyAla-LysGlyGluProGlyAlaArgGlyGluArgGly 449  
QY 3340 GAAGGAGGAGGCGAGCGCTGGCTCACCACCAAGGAGATACCTGCTCCCAAGTTTCT 3281  
DB 450 GluAlaGlySerProGly-----IleProGlyProLys----- 460  
QY 3280 ATTTTGTGTGCTCCCTGCACTGGCCCTGTCTCCATGCTAGGCTCTTTTGTTCATGGCCT 3221  
DB 460 ----- 460  
QY 3220 CCCTGAGTTTTTCCAAAGGATATCATTTGTATCCAGGTGACAGTTGAAGCAGTGGCGG 3161



```
QY 1234 GGGGATGGAGGAGGGGAAACACTCTCTGCAAGACGGGAGAGAAAGAGGCGATGTGTATG 1175
||||| |||
Db 1031 GlyAspArgGly----- 1034

QY 1174 TTTGAGGATGAGGACTTTGGCAGCGCATGATGTCTATCAGTAGATACATTGAGAATGGATCT 1115
||||| |||
Db 1035 ----- 1038

QY 1114 GTGTGTGAGTGTGAGAGAGTCTTTGGCATGTGTGCCACTGTGTGTGTCACCATGTGTGTCACCATGTATAT 1055
Db 1038 ----- 1038

QY 1054 GTGTGTATGAGGAGGAGCGGGCAGGCGTAGGACTGCAGGTGCAGCTGCTGTGTGTTC 995
||||| |||
Db 1039 -----ProGlyAla-----ProGlyAlaProGlyHisProGlyProProGly 1052

QY 994 AGTTGTGA-----CCCTAGCTTCTCTGC 974
||||| |||
Db 1052 yProValGlyProSerGlyLysSerGlyAspArgGlyGluThrGlyProAlaGlyProSe 1072

QY 973 CAGTCTTCCAGTCCCACTGCAGCTTGAGGGAGTCACAGGTGGGAGACACACATGTCA 914
||||| |||
Db 1072 rGlyAlaProGlyPro-AlaGlyAlaArgGlyAlaProGlyProGlnGlyProArgGlyA 1092

QY 913 GAGCTGGCAGGCTGAGCACAGTCAGCTCTAAGGGAGGAAGCAGGAGAGAGCGCGCGCCA 854
||||| |||
Db 1092 spLysGlyGluThrGlyGluArgGlySerAsnGly-lleLysGlyHisArgGlyPhePro 1111

QY 853 AGAACAGATCACCCATAGGAAGCAAGGTATAGGGTGTAGACCATAGGGTGGGCTGCT 794
||||| |||
Db 1112 GlyAsnProGlyProProGlySerPro-----GlyAlaAlaGlyHisGlnGly----- 1127

QY 793 TATGGGCTTAGTTTGGGTATATCTTAGGATGTATCTCTAGTGTGGGAGGTAGGG 734
||||| |||
Db 1128 -----AlaIleGlySerProGly 1133

QY 733 TGGGGTTAGATTGAAGGAGATTTAGTGCTCACTCTCAGGGAACGGTGGCAGTGTGGA--- 677
||||| |||
Db 1134 -----ProAlaGlyProArgGlyProValGlyPro 1143

QY 676 ---GGTTCAACCGTCTTGTGTAGTGGGGAAGTGTGAGCTCAATGAGGCGCTTTC 620
||||| |||
Db 1144 HisGlyProPro-----GlyLysAspGlyThrSer----- 1153

QY 619 CATCAGCTGCAAGGTGGGAGGAGGAGGAGAGGTGCATTTGACGAGTCCGCGGCTTG 560
||||| |||
Db 1154 GlyHisProGlyProIleGlyProProGlyProArgGlyAsnArgGlyGluArgGly--- 1172

QY 559 TTGGCTGATGGCACTGCAGAGTTTGGACATCCCAAGGATACCAGAGTCCAGAGCGCAGGG 500
||||| |||
Db 1173 --SerGluGlySerPro-----GlyHisProGlyGlnProGlyProProGlyProPro 1189

QY 499 TTGGGTGCTGTAGGAGGAGTGTGGGAGTGGGGAATATGCTTTTCTGGGATTTTCAGT 440
||||| |||
Db 1190 ---GlyAlaProGlyProCysCysGlyGlyGlyAlaAlaAlaIleAlaGlyValGlyGly 1208

QY 439 GAGGACCAAGGTGAGGATTCAGTGGACACTTTGGA 405
||||| |||
Db 1209 GluLysSerGly--GlyPheSerProTyTyArgGly 1219

RESULT 10
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: T30826
R:Rifkov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826
```

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2187 <YOT>

A:Cross-references: EMBL:U48363; NID:gl666688; PID:gl666689; PIDN:AAB18732.1

C:Genetics:

A:Gene: Naca

A:Map position: 10

A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3

A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding a

C:Keywords: alternative splicing; DNA binding; transcription factor

Alignment Scores: 3.3e-12 Length: 2187  
Pred. No.: 323.00 Matches: 435  
Score: 323.00  
Percent Similarity: 31.11% Conservative: 194  
Best Local Similarity: 21.51% Mismatches: 762  
Query Match: 3.42% Indels: 631  
DB: 2 Gaps: 89

US-09-931-704-3 (1-5087) x T30826 (1-2187)

```
QY 60 CTCTCCACTCCGCGAGCTCCGGGAGAGGCGGCGACCC----- 98
||||| |||
Db 105 ProProIleSerProAlaAlaLeuAlaLeuAlaSerProMetIleGlyLeuAlaGlnLys 124

QY 99 --GSCCGCGCCAGCCCGCAGCCCATGGACCTCCGAGCAGAGTTGAAACCCCAACTAGCC 155
||||| |||
Db 125 GlyAlaArgSerSerAlaProLeuSerLeuValAla----- 137

QY 156 CTGCTCTTCTATAACATGACAGCAGCGCCCATCTGATA----- 194
||||| |||
Db 138 -----LeuAlaProHisSerValGlnLysSerSerValCysPro 150

QY 195 CCTAACCCGACCAAGTCACAGCCCTCCAACTCACCCTCTGCCTGCCGACGACCTCACCACA 254
||||| |||
Db 151 ProHisProLeuThrSerProSerAlaAlaGlyAlaGluLeuGlyAlaLeuThrAla 170

QY 255 TCCTTGTGGCTCAAAACCTCAACCGCCTAAATCAACCAAAATCCCAAGTCTAAACTAATC 314
||||| |||
Db 171 SerIle----- 172

QY 315 TGAAACTTTTAAAGTAACCCAGCTCTTAAACCTAACCTAGAGCCCAATATATATCTA 374
||||| |||
Db 173 -----ProProLeuGluPro----- 177

QY 375 CCCTAGCCAAACCCCTAACTGCTTTGCCAGTCCAAAGTGTCCACTGAATCCTCACCCTGG 434
||||| |||
Db 178 -----LysThrSerThrSerGlnValPro--- 185

QY 435 TCCTCACTGAAAAATCCAGAAAAGCATATTTCCCACTGCCACATCCCTCTTACAGCA 494
||||| |||
Db 186 SerGlnGlyThrLeuAsnLeuLysGlyThrAlaProCysProProAspValValArgAla 205

QY 495 -----CCCAACCCCT---GGCTCTCGACTCTGTGTATCTCTGGATCTCCCAATCTCA 536
||||| |||
Db 206 PheProSerHisLeuGluAsnProLeuAlaSerValGlnProGlyLeuMetSerCysPro 225

QY 537 AACTCTGAGTGCCATCAGCCAAAGCCGACTCGTCAATGCGACCTCTCTCCCTCTCT 596
||||| |||
Db 226 GlnThrLeu-----Ser-AsnThrSerProValLysGlyValProIleSerSerAlaLe 243

QY 597 GTCCCCACCTTGCAGGCTGATGGAAGGCTCTATTGAAGTCCCACTTTTCCCACTTAA 656
||||| |||
Db 243 uThr-----GlnSerArgLeuSerLeuAsnLeuLysGlyProValSerProPro-- 259

QY 657 CACCAAGAACCGGGTGAACCTCCACACTGCCACCTCCCTCTGAGAGTGAGCTAAATCT 716
||||| |||
Db 260 -----AlaArgAsnThrAlaAlaProSerIleProLeuAlaProSerThrSerLe 276

QY 717 CTTTCAATCTAAACCCACCTTACACTTCCACACTCAGGAATCAGATCTCTAGATATACC 776
||||| |||
Db 276 uGlyCysHisLeuProLeuLeuHis-----HisSerSerValAspSerProIleGlnPr 294
```



```
QY 777 CAATAAAGCCATAGGAGCCGACCCCTAGTGTCTTAACCTATACCTGCTTCT 836
Db 294 oProGlyGlnSerGlyLeuAlaVal-----SerAsnPro----- 305
QY 837 ATGGGTGAGTCTGTTTGGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 896
Db 306 ----ThrSerValGlyHisSerGlyIleAlaAlaSerCysProProGluArgCysVal 324
QY 897 GCTCAGCTGCTGACATGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 956
Db 324 lProAlaLeuProSerArgLeuLeuAlaValAspSerGlyAlaAla-ProSerAsp 344
QY 957 GTGGG-----ACTGGAAGACTGGCAGGAAGCTAGGTACAACTGGAACACAGG 1004
Db 344 ysGlySerSerAlaValThrAsnGluLeuCysSerProProGlySerSerAsnVal 363
QY 1005 CAGTCCACCTGACGTCCCTAGGCTGGCCCGCTCCCTCCATGTACACATATACATGT 1064
Db 363 laGlyThrSerLeuSerProLysAlaSerLeuValProLysGlySerAsnVal 380
QY 1065 TGGCACACACACAGTGGCACATGCCAAGACTCTCTCAGCTGACACACAGATCCATTC 1124
Db 381 -----AlaL 382
QY 1125 TCAAGTATCTACTGATAGACACTCATGCTGCCAAGTCTCATCTCAAACT 1177
Db 382 euGlnProLeuValThrGlnValProAlaSerGlnLysThrGlyLeuLysGluIlePro 402
QY 1178 -----ACATGCTCTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1220
Db 402 alSerCysIleGlyAlaThrHisAlaLeuAspAsnProSerAlaIleSerValAla 422
QY 1221 CCTCTCTC-----ATCCCTCTGCTCCATCTGCTGCTGCCACCTCACCCCCAC 1271
Db 422 roAlaThrHisValProProProThrSerSerGlyLeuValSer-----SerLys 441
QY 1272 CCAGCCCAAGTGGGGACAGACACTGAGGGGCTGCCAGCTGCTTCCCGGTGGGGCCG 1331
Db 441 laSerPro-----ValThrSerLeuValProAlaAlaHisLysGlnPhePro 458
QY 1332 GCGCGGCTCATGCTTCTCG-----TCCATCTGCCACAGGGACTCGT 1376
Db 458 laProProAlaSerAlaThrLeuGlyValProValSerProLeuProAlaThrGluGly 477
QY 1377 GGGGATGTTAGCGCTGTGACGGTCTGTCGACCTCCCTGACGTGCGACGTGCA 1436
Db 478 -----LeuLysAsnLeuPro-----IleSerAlaLeu 487
QY 1437 ATCGCACAGGGACCCA-----GGGCTGGCCCTCCATCCAGAAAACCTATG 1484
Db 487 alAsnValGlyAlaProValSerProAlaGlnAlaGly-LeuProThrArgLysAsp 506
QY 1485 ACC-----TCACCGCTACCTGGAGCACCAC 1511
Db 507 ThrLeuGlnProLeuAlaProIleAlaLeuLysGluSerProSerSerGlnSerAla 526
QY 1512 TCCGCGAGCTGCTG-----GGACCTATGGA 1538
Db 527 SerLeuGluValLeuSerGluAspThrValThrLysLysThrThrGlyGlyProAla 546
QY 1539 GTATCCAGCGTAGGAATCTGGAGTTGGGAGGAGTGGAGGAGTGGGAAAGACAGTCT 1598
Db 547 ValValArgProAlaIleAlaGlyValAlaThrThrSerLeuArgAlaAspSer 565
QY 1599 AACCGTGGAGGTTCTGTAATGATGGGTGAGGAGGGGCTCTTGGCTCCACCATC 1658
Db 566 -----ProAla 568
QY 1659 CCCGTGCTGCTATCTCTGCTCCCTCTTCTAGTGGCCCGCCCACTT----- 1709
Db 569 VallieArgAlaAspSerCysValSerProAsnThrValSerGlnProLeuLysArg 588
QY 1710 -----CCCCATCCCTGGCCCCCAGGACTAGGCATGTGGGCGAGGCTCGCACCCGCT 1760
```

---

```
Db 589 ValThrAspProAlaMetaAlaProArgThr---AlaLysAsnThrAlaProSerThr 607
QY 1761 TGGCCCATTTGCCCATCTGGCT-----GCCAGCCAGCCGCGCGCTC 1802
Db 608 SerProLeuValProLeuAlaSerGluGlyCysProValAlaSerSerMetAlaLeu 627
QY 1803 CCCCTGGGGCGGGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1862
Db 628 ProGlnAsnAla---SerValSerGluThrAlaLeuAlaLeu----- 640
QY 1863 CCGGTTGGGTGGGACAGAGGGGCCACCTCCCATGCGCTTCCAGCTCCGCTCTG 1922
Db 641 -----SerProGluIlePro-----LysSerValProPhe 650
QY 1923 CCC---CCAGACCTGGGGCCCTGCTGC---TCTGGACCCAGGGCCCTCTCTCTCT 1976
Db 651 ProAspProProLeuAlaGluIleSerPheSerAsnAlaArgLysValAspAlaVal 670
QY 1977 CTCCTCCATCTAGCTGGGCTCTCTAGGGGCTCATAGGGGAGGAGGAGGAGGAGG 2036
Db 671 HisMetGluSerSerGlySerSerArgGlnGlyHisProAspAlaSerValThrAla 690
QY 2037 CAGCAGTACTGGCAGGGGTTTAGGTTGATGGAGTTATCTCTAAGATTGGGG 2096
Db 691 GlyThrValVal-----CysLeuAlaAspSerSer 700
QY 2097 GTGCTGCAGAGGTGTTTCAGAGACCCAGAGAGAGAGAGAGGAGGTGGAGAGCC 2156
Db 701 LeuAspThrSerValSerAlaSerLysGlySerAlaLeuSerGlyAlaSerSerPro 720
QY 2157 CAC----- 2159
Db 721 TyrProLeuGluValSerPheLeuProGluAlaGlyLeuAlaValGlnGlyProLys 740
QY 2160 -----CATGGGAACCGCCCTCTTCC 2183
Db 741 SerLeuAsnLysLeuSerProThrProSerSerLysGlyAlaProValProSerThr 760
QY 2184 CGTGTCTCTCTCCACATCCACACCTCTCTCTGAGCC-----GlyAlaProIleValProThrGluSer 776
Db 761 GlyAlaPro---ProSerProLys----- 796
QY 2226 GAAAGAAAAGGAAGAAGTGGCGGGAGTGGCTCCAGCCCGCAGGATACACCGAGAA 2285
Db 777 SerIleSerSerLysGlnValProAlaGluIleLeuProSerProGlnLysThrPro 796
QY 2286 ATTAGTTGTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2342
Db 797 ValThrAlaSerArgLeuIleSerAla-ValGlnSerProLysValAspProIleMet 816
QY 2343 CCAGGCTCTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2388
Db 816 AspValThrPro---ThrSerProLysLysThrSerAlaThrAlaValProLysAsp 835
QY 2389 -----CCCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2420
Db 835 rSerAlaThrLeuSerLeuLysSerValProAlaValThrSerLeuSerProProLys 855
QY 2421 CCCGAGGGTGGAAAGGGCTCTGCCCTCTTCCCTATACCATGCTGCTCTCCATAGCT 2480
Db 855 aProValAla-----ProSerAsnGluAlaThrIleVal-ProT 868
QY 2481 CCTCTGCTCTTACTCATGAGACTGCTCCATCTTCTCTCTGCAACCTGCTCTCATCA 2540
Db 868 hrGluIleProThr-----SerLeuLysAsn-AlaLeuAlaAla 880
QY 2541 GCTGAACCTCTCTTTCGGAGTGTAGTGAGTACCGCTCTCTCCCGCCCTCTCA----- 2594
Db 881 AlaThrProLysGluThrLeuAlaThrSerIleProLysValThrSerProSerPro 900
QY 2595 -----GCTGGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2630
```





Db	293	GlyGluLysGlyGluLysGlyGluGlnGlyGluProGlyLysArgGlyLysProGlyLys	302
Qy	1338	GCTCATGCTTCTCGTCCATCTCTGCCACACAGGGGACTCGTGGGGAGTGTTAGCGTGCTGT	1397
Db	303	AspGlyGluAenGly-GlnProGlyLeuProGlyLeuProGlyAspProGlyTyrPro--	321
Qy	1398	GCACGGTCTCTGGCACCTCCTGCAGTGCACAGTCTCAATCGCACAGGGACCCAGGC	1457
Db	322	-----GlyGluProGlyArgAspGlyGluLysGlyGlnLysGlyAspThrG1	337
Qy	1458	CTGGCCCTCCATCCAGAAAACCTATGACTCACCCG-----	1494
Db	337	yProGlyProProGlyLeuValleProArgProGlyThrGlyleThrleGlyG1	357
Qy	1495	-----CTACCTGGAGCACCACTCCGACGTTGGCTGGACCTATGTGA	1538
Db	357	uLysGlyAenleGlyLeuProGlyLeuProGly-----	368
Qy	1539	GTATCCACGCTAGGAATCTGGAGTTGGGA-----GGAGTGAGAGT-----	1581
Db	369	-----GluLysGlyGluArgGlyPheProGlylleGlnGlyProProGlyLe	384
Qy	1582	-----TGGGAAAGACAGTCTTAACCTGGAGGGTCTGGTAAATGATGGGGTGA	1631
Db	384	uProGlyProProGlyAlaAlaValMetGlyProProGlyProProGlyPheProGlyG1	404
Qy	1632	GGAGGGCTCTTTGGCTCCACAGCTCCCTCTGTGTCTATCTCCTGCCCTTCCCTCT	1691
Db	404	u-ArgGlyGlnLysGlyAspGluGlyProProGlylleSerleProGlyProProGlyL	424
Qy	1692	TAGTGGCCCCCCCCATTCCTCCATCCCTG-----GCCCCAGACTAGGCATGGGCAG	1745
Db	424	euAspGlyGlnProGlyAlaProGlyLeuProGlyProProGly-----	438
Qy	1746	GCCTCGCACCCGCTTGGCCCATTCGCCACTGGC-----TGCCAGCCCGCCGCC	1796
Db	439	-----ProAlaGlyProHisleProProSerAspGluLecysGluProGlyProP	456
Qy	1797	CGCTCCCTCGGGGCGGGGAAGTCTCTCTGTTTACACCGTGTGTGGTGTCTCTTG	1856
Db	456	roGlyProProGlySerProGlyAsp-----	464
Qy	1857	CGCGGGCGGGTGGGTGGGACAGAGG-----	1885
Db	465	-----LysGlyLeuGlnGlyGlnGlyValLysGlyAspLysGlyAspThrCysPheA	483
Qy	1886	-----GCCCCACTCCATGCTGGACCCAGGGGCTCCCTCGTCTGCCTCTG	1922
Db	483	snCyslleGlyThrGlylleSerGlyProProGlyGlnProGlyLeu-----	498
Qy	1923	CCCCCAGACTGGGGCCCTGCTCTGACCCAGGGGCTCCCTTCGCTGCCTCTCC	1982
Db	499	-----ProGly-----LeuProGlyProProGlySerLeuGlyPheP	511
Qy	1983	CATCCTAGCTGGCCCTCTAGGG-----GGGTCTATGGGG-----AAGGGGACTGTAGGGA	2033
Db	511	roGlyGlnLysGlyGluLysGlyGlnAlaGlyAlaThrGlyProLysGlyLeuProGlyI	531
Qy	2034	ACCCAGCAGTAGTGGCAGG-----GGGTTAGGTTGGATGGAGGTTATGCTCTAAGGAT	2090
Db	531	leProGlyAlaProGlyAlaProGlyPhePro-----GlyS	543
Qy	2091	TTGGGGTGGTCCA-----GAGGTGTTTCAGAGACCCAGGAGAGAAGAGAGG-----G	2141
Db	543	erLysGlyGluProGlyAspilleLeuThrPheProGlyMetLysGlyAspLysGlyGluL	563
Qy	2142	TTGAGGAGCGGAGGCA-----CCATGGGGAAACCGGCCCTCTTCC	2183
Db	563	euGlySerProGlyAlaProGlyLeuProGlyLeuPro-GlyThrProGlyGlnAspGly	582
Qy	2184	CGTGTCTCTTCATCTCCAGACCCCTACTCTGGAGCCAGGGAAGAAAGAGAGAGAG	2243
Db	583	LeuProGlyLeuProGlyProLysGlyGluProGlyGlyleThrPheLysGlyGluArg	602

Db	511	roGlyGlnLysGlyGluLysGlyGlnAlaGlyAlaThrGlyProLysGlyLeuProGlyI	531
Qy	2034	ACCCAGGCGAGTAGTCGACAG---GGGTGTAGGCGTGGATGAGGTTATGCTCTAAGGAT	2090
Db	531	leProGlyAlaProGlyAlaProGlyPhePro-----GlyS	543
Qy	2091	TTGGGGGTGGTCCA---GAGGTGTTTCAGAGAGCCAGGAGAAAGGAGGAGG-	2141
Db	543	erLysGlyGluProGlyAspIleLeuThrPheProGlyMetLysGlyAspLysGlyGluL	563
Qy	2142	TTGAGGAGGCCGAGGCA-----CCATGGGGAACCGCGCCCTCTTCC	2183
Db	563	euGlySerProGlyAlaProGlyLeuProGlyLeuProGlyGlnAspGly	582
Qy	2184	CGTGTCTCTTCACATCCAGACCCCTACTCTGGAGCCAGGGAAGAAAGGAGAGAG	2243
Db	583	LeuProGlyLeuProGlyProLysGlyGluProGlyGlyIleThrPheLysGlyGluArg	602

Db	543	erLysGluIupProGlyAspIleLeuThrPheProGlyMetIysGlyAspIysGlyIuL	563
Qy	2142	TTGAGGAGCGCAGGCA-----CCATGGGGAAACGGGCCCCCTCTTC	2183
Db	563	eUGlySerProGlyAlaProGlyLeuProGlyLeuProGlyGlnAspGly	582
Qy	2184	CGTGTCTCTTCACATCCACACCCCTACTCTGGAGCCAGGGAAGAAAGAGAGAAG	2243
Db	583	LeuProGlyLeuProGlyProGlyGlyGluProGlyGlyIleThrPheIysGlyGluArg	602

[illegible]

583	LeuProGlyLeuProGlyLeuProGlyLeuPro-GlyThrProGlyGlnAspGly	582
563	euiGlySerProGlyAlaProGlyLeuProGlyLeuPro-GlyThrProGlyGlnAspGly	582
Db		
Qy	2184 CGTGTTCCTTCCACATCCCNAGACCCCTACTCTGGNGCCAGGANAAGAGGAGNAAG	2243
Db	583 LeuProGlyLeuProGlyProGlyGlyGluProGlyGlyIleThrPhenylsGlyGluAsp	602

Qy	2184	CGTGTTCCTCTTCCACATCCACAGACCCCTACTCTGGAGCCAGGGAAAGAAAAAGGGAAG	224
Db	583	LeuProGlyLeuProGlyProLysGlyGlyLeuProGlyGlyLeuThrPheLysGlyGluArg	602

Db 583 LeuProGlyLeuProGlyProLysGlyGluProGlyGlyIleThrPheLysGlyGluArg 602



Db 1443 -----LeuAspGlyProProGlyProAspGlyLeuGlnGlyPro----- 1455

Qy 5042 GTCCACCCACCCAGGTTTCCT 5060  
||||| |||

Db 1456 -----ProGlyProPro 1459

RESULT 12

G86292

hypothetical protein F7H2.17 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C:Accession: G86292

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, Chan, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; ansen, N.F.; Hughes, B.; Huizlar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maifi, R. Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H. ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86292

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1006 <STO>

A:Cross-references: GB:AE005172; NID:G8927662; PIDN:AAF82153.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Alignment Scores:

Pred. No.:	5,17e-12	Length:	1006
Score:	319.50	Matches:	247
Percent Similarity:	27.35%	Conservative:	64
Best Local Similarity:	21.72%	Mismatches:	310
Query Match:	3.39%	Indels:	517
DB:	2	Gaps:	50

US-09-931-704-3 (1-5087) x G86292 (1-1006)

Qy 2151 CCGAGGCACCATGGGAACCGGCCCTCTTCCCGTTCCTCTTCCA----- 2198  
||||| |||

Db 103 ProArgProArgProArgProArgProArgProArgProArgProArgProArgProValProSer 122  
||||| |||

Qy 2199 -----CATCCACACCTACTCTGGAGCCAGGAAGAAAGGGAAGAAG 2243  
||||| |||

Db 123 ProProProProLeuHisProArgPro----- 131

Qy 2244 GTGGCGGGGAGCTGGCTCCAGCCCGGAGGATACACCGAGAAATTAGTTGCTCTGTGC 2303  
||||| |||

Db 132 -----SerProCysProProProLeuMetProSerProProProLeuVal 146  
||||| |||

Qy 2304 TTGTCAGCGGTGAACCTCCCTCGGCCCTTG-----CCTATCCAGGCCTCTCC 2354  
||||| |||

Db 147 ProSerProProProProProSerProSerProLeuValProSerProProProSerPro 166  
||||| |||

Qy 2355 CCTTGCTTCCTCTCTTCCAGTTATACATCCCTCATCCCTTCCCT----- 2405  
||||| |||

Db 167 ProProPhePhePhePheProSerProProProProValIleValPheProProLeu 186  
||||| |||

Qy 2406 -----GGGCCCCAGCGCTCCCGGAGGCTGCGAAAGGCTCTGCCCTTCCCTA 2456  
||||| |||

Db 187 ValProSerProProProProProGlyAspGlnThrGlnProProProLeu 206  
||||| |||

Qy 2457 TACCATGCTGTCTTCCATAGCCTTCTCTGCTCTACTCATGAGCTGCCTCATTTCTT 2516  
||| |||

Db 207 Trp-----LeuProPro-----ProProPheGly 214  
||| |||

Qy 2517 CCTTCTGCAACCTGCTCCTATACAGCTGAACCCCTTCTTCGGAGTGTAGTAGTACCG 2576  
||| |||

Db 215 AspGluThrProProValPheSerLeuProPro-----ProLeuAspGluProPro 231  
||| |||



709	----	-----AlaProGlyIleThrGlySerProProProLeuVal	721
4481	AGC	-----ACAGAGTGGCAACAAACCCAGAGCTGAGCATCAGG	4519
722	AlaAspValProProMetProProLeuAlaTyrPheSerProProAspIleThr	-----	739
4520	ACCTTGGCTCGAATTGTCTTCAGATATTACGGTGGCTCTTCTGCCCCTTTCACAGG	4579	
740	-----ThrGlySerProProProSerProVal	748	
4580	TATCTGTGGTTCACAGCTGGGAGGCCAACATA-----GCCACACCAAGGAT	4630	
749	PheLeuLeuProProProLeuAspArgSerThrLeuThrProProAlaAlaProValAsp	768	
4631	TTCTCTGAAAGTTTACAATGCAGTAGCAATTTGGGGTGTAGGGTGCGAGCTCCCAAGGCC	4690	
769	-----AsnLeuProProVal	773	
4691	CTG-----CCCCCAGCCCCACCATCATGACTCTAAGTGTGTATTATATAT	4741	
774	IleIleThrGlySerProProProValAsnAsn-----	784	
4742	TATTTATTGGAGATGTTATTATTATAGATGATATTATTTCAGAAATTTCTATTCTTGTTAT	4801	
784	-----	784	
4802	TAACAAATAAAATGCTTGCCTCCAGAACTTAGTC-----	4834	
785	-----LeuProProAspIleValIleGlyGlnProProProAspVal	799	
4835	-----TCTTTGCCAGCCTCACCCCTCTCTGGTGCTCATCAGACTCTTT	4876	
800	ThrIleGluProProIleAspGlnSerThrLeuThrProProProValIleProValThrIle	819	
4877	GCCACCCCTGGCTCCACTCCCTGTGGCTCTGTGTGGAGTGCACAGAGCTCTGGGAAG	4936	
819	uProPro-----ProValGlnAspLeuProSerIle-----	829	
4937	AGGCCCTCTTCTCCCGCACTGGGGCGATGGCGCACCT-----CAGACTTACCCAGTCT	4993	
830	-----LeuProProProAlaAspGluLeuProProProValGlnGluPheProProIle	847	
4994	GCTGCCACCAACCAACC-----CCTTTGATCCCTCAGTCTCCCCACACAGCTTCTGTC	5044	
847	eLeuProProProValGlnAspPheProProIleLeuAlaProProAlaAspGluPhe-P	867	
5045	CACCCAGGTTTCCTTCACCCACCACTTTGGCTAAAGTCTTCCT	5085	
867	roPro-----AsnLeuProProProValLeuGluPhePro	878	

RESULT 13  
S06733 hydroxyproline-rich glycoprotein precursor - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 21-Jul-2000  
C:Accession: S06733  
R:Keller, B.; Lamb, C.J.  
Genes Dev. 3, 1639-1646, 1989  
A:Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene  
A:Reference number: S06733; MUID:90128263; PMID:2612909  
A:Accession: S06733  
A:Molecule type: DNA  
A:Residues: 1-620 <KEL>  
A:Cross-references: EMBL:X13895; NID:g19866; PIDN:CAA32090.1; PID:g19867  
C:Superfamily: Hydroxyproline-rich glycoprotein  
C:Keywords: glycoprotein

Alignment Scores:	
Pred. No.:	1.02e-11
Score:	314.50
Length:	620
Matches:	217
Conservative:	42
Mismatches:	268
Indels:	326

DB:	2	Gaps:	44
US-09-931-704-3	(1-5087) x 806733	(1-620)	
QY	1658	CCCCCTGCTGGTCTATCTCTCTGCGCCCTTCCCTCTTAGGT---	GGCCCCCCCACTTCCCCA 1714
Db	36	ProProValThrSerGlnProProSerSerIleGlyLeuSerProSerAlaPro	55
QY	1715	TCCCTGGCCCCAGGACTAGGCATGTGGCAGGCCTCGCACCCGCTTGGCCCATTCGCCCC	1774
Db	56	ThrThrThrProProSer-----ArgGlyHisValProSerProArgHisAlaPro	72
QY	1775	ACTGGCTGCCAGCCAGCGCCGCGCTCCCTCGGGGGCCGGGAAGTCTCTCTGTTTA	1834
Db	73	-----ProArgHisAlaTyrProProProSerHisGlyHisLeu-----	85
QY	1835	CACCGTGTGTGGTGTCTCTTGGCGGGGGGTGGGTGGG-----	1876
Db	86	-----ProProSerValGlyGlyProProProHisArgGly	97
QY	1877	-----GACAGAGG---GCCCCACTCCCATGCTCCGTCGTCACG---TCGCC	1918
Db	98	HisLeuProProSerArgGlyPheAsnProProProSerProValIle-SerProSerHi	117
QY	1919	TCTGCCCCAGACTGGGG-----CCCTGTGCTCTGTACCCAGGGGCTCCCTTCCGT	1972
Db	117	sProProSerTyrGlyAlaProProProSerHisGlyProGlyHisLeuProSer--	136
QY	1973	CTGCCTCTCCCATCTAGCTGGGCTCTAGGGGGGTCAATGGGGGAAGGGACTGTAGGG	2032
Db	137	-----HisGlyGlnArg-----	140
QY	2033	AACCCAGCAGTAGTGGCAGGGGTTTAGGGTGTGGATGGAGGTATGCTGAAGGATTT	2092
Db	140	-----	140
QY	2093	GGGGTGGTCCAGAGGTGTTTCAGAGAGCCCGAGAGAG-----AAGGAAGGAGG	2140
Db	141	-----ProProSerProSerHisGlyHisAlaProProSerGlyGI	154
QY	2141	GTGGAGGAGCGGAGGACCATGGGGAAACCGGCCCTCTTCCCGTGTCTCTTCCA--	2198
Db	154	yHisThrProProArg-----GlyGlnHisProProSerHisArgArgProSerPro	172
QY	2199	-----CATCCCAGACCTACTCTGGAGCCAGGGAAGAAAGGAAGGTGCGGG	2251
Db	172	oSerArgHisGlyHisPro-----	178
QY	2252	GGAGCTGGCTCCAGCCCGCAGGATACCGAGGAAATTAGTTGTCTCTGTGTTGTCAGC	2311
Db	179	-----ProProThrTyr-----	183
QY	2312	GTGTGAACCTCCCTCGGCGCCTTG-----CCTATCCAGCGCTCTCC-----CCTTG	2359
Db	184	-AlaGlnProProProThrProProThrProProSerProGlnValGlnProProPro	203
QY	2360	CTTCTCCCTTTCCTCCAGTATACATCTCCCTCATCCCTTTCCT-----GGGCC	2410
Db	203	rTyrSerProProProProThrHisValGlnProThrProSerProSerArgGlyHi	223
QY	2411	CCAGCCCTCCCCGAGGGTTGAAAGGGCTCTGCCCTCTTCCCTATACCATGCTGCTT	2470
Db	223	sGlnProGlnProProThrHisArgHisAlaProProThrHisArgHisAlaProPro	243
QY	2471	CCATAGCTTCT	2530
Db	243	rHisGlnProSerPro-----LeuArgHisLeuProPro-----	254
QY	2531	GCTCCATCAGCTGAACCTTCTTTCGGAGGTGTAGTCAGTACCGTCTCTCCCGAGCCC	2590
Db	255	-----SerProArgGlnProGlnProProThrTy	265
QY	2591	CTCAGCTGGTGGGCTGGGTGTGTGAGCGGCAAAATGGGGCTCTGGTTCAAATGGGCAC	2650



```
Db 265 rSerProProProAlaTyrAlaGlnSerProGlnProSerProThrTyrSerPro-- 284
QY 2651 CTATCTCTCTCTGTTCTGTGCGAGAAACCTTGTCTTCACTCCACTGCGCTCTCTAG 2710
Db 285 -----ProProProThrTyrSerProProProProSe 295
QY 2711 TTCCCGACCCCTTTTCTCTCTGCTTCTCCCTGCGCAAAATTTCTCCAGAGGTGCTTACA 2770
Db 295 r-----ProIleTyrSerPro-----ProProProAlaTyrSe 306
QY 2771 CCCTCTGCTCCACTTCTCTCCACCACCTCACTTCTTAAACCCCTGCAATCTGCTTCC 2830
Db 306 rProSerProProProThrProThrPheSerProProProProAlaTyrSerPr 326
QY 2831 AGGCCCCAGCAATGGTCTCTCCAAAGTGTGTCAGGCACCTCTTGCACAGCCGACAGTG 2890
Db 326 oProProThr-----TyrSerProPro----- 333
QY 2891 TTTTGAAGGCTCATCTCTCTGCTGTCTGTTTGTGAGCCACACTGCTGAGCGCTGCTGCC 2950
Db 334 -----ProPr 335
QY 2951 TTCTCGAACTCTCTTCTCTGCTGTCTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCT 3007
Db 335 oThrTyrLeuProLeuProSerProIleTyrSerProProProProValTyrSerPr 355
QY 3008 AGCTCTCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3067
Db 355 oProProProProSer-----TyrSerProProProProThrTy 368
QY 3068 TTGCCCCACCCAGCAATCAGCAGCTCTCTCTCTGAGCGTCTTGTGCTCTCTCTCTCTCT 3127
Db 368 rLeuProProProProProSerPro-----ProProProSe 381
QY 3128 TTTTCTTACGGCTCTTCCA-----TTGAGAGCTCACACCGCCGCTCTCTCAACTGCTCA 3181
Db 381 rPheSerProPro--ProProThrTyrGluGlnSerProProProProAlaTyrSer 400
QY 3182 CCTGCATACAAATGATATCTTATTTGAAAACTCAGGAGGCGCATGACAAAGAGCT 3241
Db 401 ProProLeuProAlaProProProThrTyr-----SerProProProProThrTyrSerPro 418
QY 3242 AGCATGGAGACA-----GGGCGAGTGTGAGCGGCTCAGGGGACACAAAAATAGAACTTT 3289
Db 419 ProProProThrTyrAlaGlnProProProLeu----- 429
QY 3290 GGGAGCAGGTATCTCTTGTGTTGAGCGGCTCTGCGCTCTCTCTCTCTCTCTCTCTCTCT 3340
Db 430 -----ProProThr--TyrSerProProProProAlaTyrSe 441
QY 3341 CCCATCACCTCTCTCTTTTTCACAGCTGAACCTGCGGCCCTCTTCAACGAGCCAGAC 3400
Db 441 rProProProProPro-----ThrTyrSerProProProProThrTyrSe 456
QY 3401 TTCAACCTCTCCGCTGCGGCGCAGAGACTCTGCGCGGCGCCTGTTGACTTGGAGGTG 3460
Db 456 rProProProProAlaTyrAlaGlnProProProProPro----- 470
QY 3461 TGGCGAAGCCTCAATGACAACTGCGGTGACCCAGACTAGAGGCTTACAGCCACTTT 3520
Db 471 -----ThrTyrSerProProProProAlaTyrSerPro--ProProProS 485
QY 3521 CTGTGTTACTTCTGCGGCTCAACCGTCAAGGCTGCCACTGCTGCTGCGCGCAGCTG 3580
Db 485 erProIleTyrSerProProProProGln-----ValGlnProL 498
QY 3581 GCCACTTCTGACACAGGCTTCCAGGCTGCTGCGGAGCATTCGGGGGTATGCGCAGCT 3640
Db 498 euProProThrPheSerProProProProProArgIleHis----- 511
QY 3641 CTGGGCTACCCACTGCCC-----CAGCGCTGCTCTGGGACT----- 3676
```

```
Db 512 -----LeuProPro--ProProHisArgGlnProArgProProThrProThrTyrGlyGln 529
QY 3677 -----GAACCCACTGACTCTCTGGCCCT-----GCCACAGTGACTTCTCTC 3718
Db 530 ProProSerProProThrPheSerProProProProArgGlnIleHisSer--ProProPr 549
QY 3719 CAGAAGATGACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3778
Db 549 oProHisTrpGlnPro-----ArgThrProProThrTyrGlyGlnProPr 565
QY 3779 AAGACTTCAACCGCTCAAGAGAGTGCAGCTCCAGCAGCTCCAGCTGCACTCACCTGCAC 3838
Db 565 oSerProProThrPheSerAlaPro-----ProProArgGlnIleHisSerPro----- 581
QY 3839 CTGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3898
Db 582 -----ProProHisArgGln--ProArgProProT 592
QY 3899 CTCCACTTGTGAGAGCCGCTGTATGCAACCTGTGAGCCAGGAGACAGAAAGC 3958
Db 592 hrProThrTyrGlyGlnPro--ProSerProProThrThrTyrSerPro----- 607
QY 3959 TGTGAGCTCTGGCCCTTCTCTGAGCCGCTGCTGATCGATCAGCCCTGCTCTCC 4018
Db 608 -----Pro 608
QY 4019 TCCCACCTCCCAAGGTCTA 4039
Db 609 SerProProProTyrGlyLeu 615
```

## RESULT 14

B35098  
MHC class III histocompatibility antigen HLA-B-associated protein 2 [imported] - human  
C;Species: Homo sapiens (man)  
C;Date: 10-Aug-1990 #sequence\_revision 06-Nov-1992 #text\_change 24-Aug-2001  
C;Accession: B35098  
R;Banerji, J.; Sands, J.; Strominger, J.L.; Spies, T.  
Proc. Natl. Acad. Sci. U.S.A. 87, 2374-2378, 1990  
A;Title: A gene pair from the human major histocompatibility complex encodes large protein  
A;Reference number: A35098; MUID:90192810; PMID:2156268  
A;Accession: B35098  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-2142 <BAN>  
A;Cross-references: GB:M33509; NID:gl79338; PIDN:AAA3585.1; PID:gl79339; GB:M31293  
A;Note: the authors translated the codon AGT for residue 97 as Gly  
C;Superfamily: collagen alpha 1(IV) chain

Alignment Scores:  
Pred. No.: 1.28e-11 Length: 2142  
Score: 313.50 Matches: 477  
Percent Similarity: 29.88% Conservative: 178  
Best Local Similarity: 21.76% Mismatches: 670  
Query Match: 3.32% Indels: 869  
DB: 2 Gaps: 119

US-09-931-704-3 (1-5087) x B35098 (1-2142)

```
QY 472 TGCCACATCCCTCTTACAGCACCCACCTGGGCTCTGGACTCCTGGTATCTCTGGAT 531
Db 39 CysProSerPro-----TrpPro 44
QY 532 GTCCAACTCTGCGAG---TGCCATCAGCAACAAGCCGAGCTCGTCAAAATGACCTCTCT 588
Db 45 AlaGluSerArgGluSerCysHisCysProAla-----TyrArgPro 58
QY 589 CCCTTCTCTGCCACCTTGCAGGCTGATGGAAGGCTCATTAAGTCCAACTTTTCC 648
Db 59 ProAlaAsnLeuProSerLeuLysAlaGluAsnLysGly-----Asn 72
QY 649 CCACCTACACC-----AAGAACGGGTGAACCTCCACACTGCCACCTGCTCC 696
Db 73 AspProAsnValSerLeuValProLysAspGlyThrGlyTrp----- 86
```

QY	697	TGAGAGTGAGCACTAAATCTCCTTCAATCTAACTCAACCCACCCCTACACTTCCCACTCTCAGGA	756
Db	87	 -----AlaSerLysGlnGluGlnSerAsp	94
QY	757	ATCACATCTAGAAATATACCCAAACTAAGCCCACTAAGGCAGCGCCGACCTAGTGGTCT	816
Db	95	 -----ProLysSerSerAspAlaSerThrAlaGlnProProGluSer	108
QY	817	AAACCTATACCTTGGTCTTCTATGGGTGAGTCTGTCTTGGCGCGCGCTCTCTCTCTGTT	876
Db	109	 -----AlaSerGln--ThrProAlaSer	119
QY	877	CTTCCCTTAGAGTGAAGTGTGCTCAGCTGCCAGCTCTGACATGTGTGTCTCCACCCCT	936
Db	119	 -----GlnProLysArgProProAlaAla	130
QY	937	CTGACTCCCTCAAGCTGCAGTGGGAGCTGCAAGACTGGCAGGAAGCTAGGGTA	989
Db	130	 -----LysThrProLeuValProSerGlyValLysSerThrAlaGlnAlaSerValThrHisG	150
QY	989	-----	989
Db	150	lyAlaHisGlyAspGlyGlyArgAlaSerSerLeuLeuSerArgPheSerArgGluGluP	170
QY	990	-----CAACTGGAAACAGGCAGGTGCGACC	1014
Db	170	heProThrLeuGlnAlaAlaGlyAspGlnAspLysAlaAlaLysGluArgGluSerAlaG	190
QY	1015	TGCAGTCCCTAGCCCTGGCCCGCTCCCTCCATGTACACATATACATGTGGCACACAC	1074
Db	190	lunGlnSerSerGlyProGlyProSerLeuArgProGlnAsnSerThrThrTirpArgSpG	210
QY	1075	ACAGTGGCACACATGCCAAGACTCTCTCAGCT--GACACACAGATCCCATTTCTCAAGTA	1131
Db	210	lyGlyGly--ArgGlyProAspGluLeuGluGlyProAspSerLysLeuHis--HisGly	228
QY	1132	TCTACTGATAGACACTCATGCGTGGCCAGTCCCTCATCTCTCAACATACACATGCCTCTCT	1191
Db	229	HisAspProArgGlyGlyLeuGlnProSerGlyProProGln-----	242
QY	1192	TTCTCTCCCGTCTTGCAGGAGTGTTCCTCTCTCCATCCCTCTCGCTCCCATCTGGT	1251
Db	243	PheProProTyrArgGlyMetMetProProPheMetTyrProProTyrLeuProPhe---	261
QY	1252	GTCCACACCTCACCCGCCACCCAGCCCAAGTGGGGACAGACACTGAGGGGTGCCAGC	1311
Db	262	ProPro--ProTyrGlyProGlnGlyProTyrArgTyrProThrProAsp--GlyProSerA	281
QY	1312	TGCTTCCCGTGTG-----GGCCCG-----	1331
Db	281	rgPheProArgValAlaGlyProArgGlySerGlyProProMetArgLeuValGluProV	301
QY	1332	---GGCCGC-----	1337
Db	301	aiGlyArgProSerIleLeuLysGluAspAsnLeuLysGluPheAspGlnLeuAspGlnG	321
QY	1338	-----GCTCATGCT-----	1346
Db	321	luAsnAspAspGlyTirpAlaGlyAlaHisGluGluValAspTyrThrGluLysLeuLysP	341
QY	1347	-----TCTCGTCCATCTCTGCCACAGGGGACTCGGGGGGATGTTAGC	1389
Db	341	heSerAspGluAspGlyArgAspSerAspGluGluGlyAlaGluGlyHis-----A	359
QY	1390	GTGCTGTGCACGGTCTCTGGCACCTCCCTGCGAGTCCGACTCTCAATCGCACAGGGGA	1449
Db	359	rgAspSerGlnSerAlaSerGlyGluGluArgProProGlu-----AlaaspGlyL	376
QY	1450	CCGAGGCTGGCCCT-----	1470
Db	376	ysLysGlyAsnSerProAsnSerGluProProThrProLysThrAlaTirpAlaGluThrS	396

QY	1471	CGAGAAACCTATGACCTCACCGGCTACTCGAGGACCAACCTCCGCAGCTT-----	1521
DB	396	erArgProProGluThrGluProGly- ProProAlaProLysProProLeuProProGly 415	
QY	1522	-----GGCTGGGACCTATCTGAGTATCCAGCG-----TAGGAATCTGGG 1560	
DB	416	AspTyrProAspArgGlyGlyProProCysLysProProAlaProGluaspGluaspGlu 435	
QY	1561	AGTTGGGGAGGAGTGAGG-----AGTTGGGGAAAGACAGTCTCTAACCTGGGAGGTTCT 1614	
DB	436	AlaTtpArgGlnArgArgLysGlnSerSerGluIleSerLeuAlaValGluArgAla 455	
QY	1615	GCTAAATGATGGGCTGAGAG-----	1635
DB	456	ArgArgArgArgGluGluGluArgMetGlnGluArgAlaAlaCysAla 475	
QY	1636	-----GGGCTCTTTGGCTCCACCACG-----	1656
DB	476	GluLysLeuLysArgLeuAspGluLysPheGlyAlaProAspLysArgLeuLysAlaGlu 495	
QY	1657	-----TCCCTCTGTCTGTCTATCTCTCTGCTTCCCTCTTCTAGTGGCCCCCCCCACTTC 1710	
DB	496	ProAlaAlaProProAlaAlaProSerThrProAlaPro-----ProProAlaVal 512	
QY	1711	CCC-----ATCCCTGGCCCCAGGACTAGGCATGTGGCAGGCCTCGCACCCCTTGGC 1764	
DB	513	ProLysGluLeuProAlaProProAla-----ProProAlaSerAla 527	
QY	1765	CCATTGCCCACTGGCTGCCACCCAGCCGCCCGCTCCCTCTGGGGCGGGGAAAGTCT 1824	
DB	528	ProThrProGluThrGluProGluGluProAlaGlnAlaPro-----ProAlaGlnSer 545	
QY	1825	CCTCTGTTTACACCGTGTTGTGCTCTCTTCCGGCGGGCGGGTT-----GGTGGGGA 1878	
DB	546	-----ThrProThrProGlyValAlaAlaAlaProThrLeuValSerGlyGlyGly 562	
QY	1879	CAGAGGGGCCCACTCCCATCCCTGC-----GTTCGA 1911	
DB	563	SerThrSerSerThrSerSerglySerPheGluAlaSerProValGluProGlnLeuPro 582	
QY	1912	GCTCGCTCTGCCCCAGACCTGGGGCCCTGTCTCTCGACCCAGGGGCTCCCTCCG 1971	
DB	583	SerLysGluGlyProGluPro-----ProGluGluValProPro-----Pro 596	
QY	1972	TCTGCTCTCTCCATCTAGCTGGGCTCTTAGGGGGGTTCATGGGGGAAGGGACTGTAGG 2031	
DB	597	ThrThrProProValProLysValGluProLysGly---AspGlyIleGlyProThrArg 615	
QY	2032	GAACCCAGGACGTAGTGGCAGGGGTTTAGGGTGTGGATGGAGTTATGCTGTAAGGATT 2091	
DB	616	GlnProProSer-----GlnGlyLeu-----GlyTyrProLysTyrGln 628	
QY	2092	TGGGGGTGGTCCAGAGGTGTTCCAGAGGCCAG-----GAGAGAAGG 2133	
DB	629	LysSerLeuProProArgPheGlnArgGlnGlnGlnGlnLeuLeuLysGlnGlnGln 648	
QY	2134	AAGAGGGTTGGAGGCGCGAGCCACCATGGGGAACCGGCCCTCTT----- 2181	
DB	649	GlnHisGlnTrpGlnGlnHis---GlnGlnGlySerAlaProProThrProValProPro 667	
QY	2182	-----CCCGTGTCTCTTCCATCTCCAGACCTACTCTGGAGCCAGGGAAA 2229	
DB	668	SerProProGlnProValThrLeuGlyAlaValProAlaProGlnAlaProProProPro 687	
QY	2230	GAAGAAG-----GGAAAGAGGTGGCGGGGAGCTGGCTCCAGGCCCAAGATT--- 2274	
DB	688	ProLysAlaLeuTyrProGlyAlaLeuGlyArgProProProMetProMetAsnPhe 707	
QY	2275	ACACCGAGGAATATGTTGTCTCTGTGCTTCTACAGCTGTGAACCTCCCC-----CTG 2328	
DB	708	AspProArgTrpMetMet-----IleProProTyrVal 718	
QY	2329	GGCCCTTGCCATCCAGGCGCTCTCCCTTGCTTCTCCCTCTCTTTTCCCAAGTTATACATCT 2388	



QY 3649 CCACCTGCCCGCCGCTGCTGGGACTGAAACCCACTTGGACTCTGGCTGCTGCCCTGCCA--- 3705  
Db 1407 ArgThrGlyProGlyArgGly-----AspLysArgSerTrpProSerProLys 1422  
QY 3706 ---CAGTGAATCTCTCCAGAAAT-----GGACGACTTCTGGCTGCTGAAGGAGCT 3753  
Db 1423 AsnArgSerArgProGluGluArgProGlyLeuProLeuProProProPro 1442  
QY 3754 GCAGACTGCTGTG---GGCTCGGCCAA-----GGACTTCAA 3789  
Db 1443 SerSerSerAlaValPheArgLeuAspGlnValIleHisSerAsnProAlaGlyIleGln 1462  
QY 3790 CCGGCTCAAGAAAGATGACGCTCCAGCAGCTCCAGCAGCTCCAGCTGGGGGCTCA 3849  
Db 1463 GlnAlaLeuAlaGln---LeuSerSerArgGlnGlySerValThrAlaProGlyGlyHis 1481  
QY 3850 TG-----GCTTCT 3857  
Db 1482 ProArgHisLysProGlyProProGlnAlaProGlnGlyProSerProArgProProThr 1501  
QY 3858 GACTTCTGACTTCTCTCTCTGCTCCCTCCCTTCAACCCCTGCTCCCACTTTGTGAGACC 3917  
Db 1502 ArgTyGluProGlnArgValAsnSerGlyLeuSerSerAspProHisPheGluGluPro 1521  
QY 3918 AGCCCTGTATGCCAACACCTGTTGAGCCAGGAGACAGAAAGCTGTGAGCCTCTGGCCCTTT 3977  
Db 1522 GlyProMetValArgGlyValGlyGlyThrProArgAspSerAlaGlyValSerProPhe 1541  
QY 3978 CTTGGACCGCTGGCGT-----GTGATCGCATGACC 4010  
Db 1542 Pro---ProLysArgArgGluArgProProArgLysProGluLeuLeuGlnGluSer 1560  
QY 4011 CTGCTCTCTCCACCTCCCAAGGCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTG 4070  
Db 1561 LeuProProHisSerSerGlyPheLeuGlySerLysProGluGlyProGlyProGln 1580  
QY 4071 TCCTGTCTCTTCTACAGGAAGTCATGCTCGAGGGAGTGTGAAGTGGTTCAGGTG--- 4127  
Db 1581 AlaGluSerArgAspThrGlyThrGluAlaLeuThrProHisIleTrpAsnArgLeuHis 1600  
QY 4128 ---GTGCAGAGCGCTCATGCCCTCTGCTCTCTGCTGCTACCA----- 4166  
Db 1601 ThrAlaThrSerArgLysSerTyArgProThrSerMetGluProTrpMetGluProLeu 1620  
QY 4166 ----- 4166  
Db 1621 SerProPheGluAspValAlaGlyThrGluMetSerGlnSerAspSerGlyValAspLeu 1640  
QY 4167 -----CTTGGCCAGTGC----- 4178  
Db 1641 SerGlyAspSerGlnValSerSerGlyProCysSerGlnArgSerSerProAspGlyGly 1660  
QY 4179 -----CCACCAGCCCTCAGTGGCAGCATCTGG-----AGG 4210  
Db 1661 LeuLysGlyAlaAlaGluGlyProProLysArgProGlyGlySerProLeuAsnAla 1680  
QY 4211 GCAGGGGTTGAGGGGCA-----CCACCACACATCCCTTTC 4246  
Db 1681 ValProCysGluGlyProProGlySerGluProProArgArgProProProAlaProHis 1700  
QY 4247 TGGGGTGAAGCCCTTGGCTGGCCCACTCTCTTGGATGGGTGTCTCCCTATCCCCA 4306  
Db 1701 AspGlyAspArgLysGluLeuProArgGluGlnProLeuProProGlyProLysGlyThr 1720  
QY 4307 AATCACTCTATACATCAATTCAGGAACAAACATGTTGGTCAATTCACACAAAAGAGA 4366  
Db 1721 GluArgSerGlnArgThrAspArgGlyThrGluProGlyProLysArgProSerHisArg 1740  
QY 4367 TGAGATTAACAGTCAGGGTGGG-----GTCTGCATT 4399  
Db 1741 ProGlyProProValGlnPheGlyThrSerAspLysAspSerAspLeuArgLeuValVal 1760

QY 4400 GGAGGTGCCCTATAAACACAGAGAGAAATACTGAAAGCACAGAGGCGGACAGACCAG 4459  
Db 1761 GlyAspSerLeuLysAlaGluLysGluLeuThrAlaSerValThrGluAlaIleProVal 1780  
QY 4460 ACCAGA-----CCCAGGAGTCTCCAAAGCACAGAGTGGCCAAACAAA--- 4501  
Db 1781 SerArgAspTrpGluLeuLeuProSerAlaAlaAlaSerAlaGluProGlnSerLysAsn 1800  
QY 4502 -----CCCGAGCTGAGCATC-----AGGACCTTGCCTCGA 4531  
Db 1801 LeuAspSerGlyHisCysValProGluProSerSerSerGlyGlnArgLeuTrpProGlu 1820  
QY 4532 ATTGTCTTCCAGTATTACGGTGCCTTCTCTGCCCCCTTCCAGGCTATCTGTGGTT 4591  
Db 1821 ValPheTyGlySerAlaGlyProSerSer-SerGlnIleSerGlyGlySerHisGlyLe 1840  
QY 4592 G-----CCAGGCTGG-----GGAGGGCAACCATATGCCACAC---CACAG 4627  
Db 1840 uSerIleThrSerLysGlnTrpArgLeuArgProGlyThrProSerLeuHisProTyAr 1860  
QY 4628 GATTTCTCTGAAAGTTTACAATCAGCAGTAGCATTTTGGGTGTAGGGTGGCAGCTCCCAAG 4687  
Db 1860 G-----SerGlnProLe 1864  
QY 4688 GCCCTGCCCC-----CCAGCCCCCCTCATGACTCTAAGTGTGTGTTATTAATATT 4741  
Db 1864 uTyLeuProProGlyProAlaProProSer-----AlaLeuLeuSerG 1879  
QY 4742 TATTTATTTGGAGATGTTATTTATTTAGATATTTATTTGCAGAAATTTCTATTTCTGTAT 4801  
Db 1879 yValAlaLeuLysGlyGlnPhe-LeuAsp----- 1888  
QY 4802 TAACAAATAAATGCTTCCCCAGAGACTTAGTCTCTTTGCCCGCAGC-----CTCACC 4852  
Db 1889 --PheSerThrMetGlnAlaThrGluLeuGly-LysLeuProAlaGlyGlyValLeuTy 1907  
QY 4853 CTCTCTGTGTCTATCAGACTCTTCCACCCCTGGCTGCCACTCCCTGTGTGCT----- 4907  
Db 1908 ProProProSer-PheLeuTySerProAlaPheCysProSerPro---LeuProAspTh 1926  
QY 4908 ---CTGTGTGAGTGCACAGAGCTCTGGGA-----AGAGCCCTCT 4945  
Db 1926 rSerLeuLeuGlnValArgGlnAspLeuProSerProSerAspPheTySerThrProLe 1946  
QY 4946 TCCTCCC-----CGCACTGGGGCGATG-----GGCGCAGCTCAGACTTACCCACTGCT 4993  
Db 1946 uGlnProGlyGlyGlnSerGlyPheLeuProSerGlyAlaProAlaGlnGlnMetLeuLe 1966  
QY 4994 G-----CTGCCA-----CCACCAACCCCTTG 5014  
Db 1966 uProMetValAspSerGlnLeuProValValAsnPheGlySerLeuProProAlaProPr 1986  
QY 5015 ATCCCTCAGTCTCCACACAGCTTCTGTCCACCCAGGTTTCCCTCACCACCTTTCG 5074  
Db 1986 oProAlaProProLeuSerLeuLeu-ProValGlyProAlaLeuGlnProProSerL 2006  
QY 5075 TAAGTCTT 5082  
Db 2006 euAlaVal 2008  
RESULT 15  
S50832  
atrophin-1 - human  
C;Species: Homo sapiens (man)  
C;Date: 14-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 07-May-1999  
C;Accession: S50832  
R;Nagatani, S.; Yanagisawa, H.; Ohsaki, E.; Shirayama, T.; Tadokoro, K.; Inoue, T.;  
Nature Genet. 8, 177-181, 1994  
A;Title: Structure and expression of the gene responsible for the triplet repeat disc  
A;Reference number: S50832; MUID:95144175; PMID:7842016  
A;Accession: S50832  
A;Status: preliminary  
A;Molecule type: mRNA

A;Residues: 1-1184 <NAG>  
A;Cross-references: EMBL:D31840

C;Genetics:  
A;Gene: GDB:DRPLA; B37

A;Cross-references: GDB:270336; OMIM:125370  
A;Map position: 12p13.31-12p13.3112p-12p

## Alignment Scores:

Pred. No.: 1.32e-11 Length: 1184  
Score: 313.00 Matches: 317  
Percent Similarity: 29.17% Conservative: 94  
Best Local Similarity: 22.50% Mismatches: 416  
Query Match: 3.32% Indels: 583  
DB: 2 Gaps: 71

US-09-931-704-3 (1-5087) x S50832 (1-1184)

Qy 1256 CACCTCACCC-----CGACCAGCCCAAGGTGGGACAGACACCTGAGGGCTGCCA 1309  
Db 160 HisProProLeuPheProSerPro-----GlnProProAspSerThrPro 176  
Qy 1310 GCTGCTCCCGTGGGCCCGCGCTCATGCTTCTCG-----TCCATCCTG 1360  
Db 177 ArgGln-----ProGluAlaSerPheGluProHisProSerValThr 190  
Qy 1361 CCCACAGGGGACTCGTGGGGGATGTAGCGTGCCTGTGCACGGTGCTCTGGCACCTCCCT 1420  
Db 191 ProThrGly-----TyrHisAlaPro 197  
Qy 1421 GCATGGCAGCTCAATCGACAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACC 1480  
Db 198 MetGluProProThrSerArgMetPheGlnAlaProProGly-AlaProProHisPr 217  
Qy 1481 TATGACCTCACCC--GCTACTCGAGCACCACTCCGACGCTGGCTGGGACCTATG-- 1535  
Db 217 oGlnLeuTyrProGlyGlyThrGlyValLeuSerGlyProMetGlyProLysG1 237  
Qy 1536 -----TGAGTATCCAGCGTAGGAATCTGGGAGTTGGGAGGAGTAGAGGATTGGGAA 1588  
Db 237 yGlyGlyAlaAlaSerSerValGly----- 245  
Qy 1589 AGACAGCTCAACCGTGGAGGTCTGTGTAATGATGGGTGAGGAGGGCTCTTTGGCT 1648  
Db 246 ----GlyProAsn-----GlyGlyLysGlnHisPr 254  
Qy 1649 CCCACCACTCCC-CCTGTCTGTCTATCTCTGCTCCCTTCCCTCTTAGTGGCCCCCACC 1707  
Db 254 oProProThrThrProLys-----SerValSerSerGlyAlaSerGlyAlaProProTh 273  
Qy 1708 TTCCCTCATCCCTGGCCCCCAGGACTAGGCATGTGGSCAGGCCTCGCACCCGCTTGGCCCA 1767  
Db 273 rLysProProThrThrProValGlyGly----- 282  
Qy 1768 TTGCCCACTGGTCCAGCCAGCCGCGCTCCCTCCCTGGGGCCGGGAAGTCTCTCT 1827  
Db 283 -----GlyAsnLeuProSerAlaProProAlaAsn----- 293  
Qy 1828 CTGTTTACACCGTGTGGGTCTCTTGGCGGGCGGGTTGGTGGGGACAGAGGGGC 1887  
Db 294 ----PheProHisValThrProAsnLeu----- 301  
Qy 1888 CCCACCTCCCATGCTGGTTCACG-----CTCGCTCTGCCCCCAGACCTGGGGC 1938  
Db 302 ----ProProProAlaLeuAlaGlyProLeuAsnAsnAlaSerAlaSerProProGly- 319  
Qy 1939 CTGCTGCTCTGGACCCAGGGCCCTCCCTCGCTGCTGCTCCATCTAGCTGGGCT 1998  
Db 320 -----LeuGlyAlaGlnProLeuProGlyHisLeuProSerProTyrAlaMetGlyG1 337  
Qy 1999 CTTAGGGGGTTCATGGGGGAAGGGGACTGTAGGGAAACCAGGCAGTAGTGGCAGGGGTT 2058  
Db 337 n----- 3125

Qy 2059 TAGGGTGTGATGAGAGGTTATGCTGAAGATTTGGGGTGGTCCAGAGGTTCAGAGA 2118  
Db 338 -----GlyMetGlyLeuPro----- 343  
Qy 2119 GCCCAGGAGAGAGGAGGGTTGGAGGAGCCGAGGACCACTGGGAAACCGCCCTCT 2178  
Db 344 -Pro-GlyProGluLysGlyProThrLeuAlaProSerProHisSerLeuProProAlaS 363  
Qy 2179 CTTCCCGTGTCTCTTCCACATCCACAG-----CCCTACTCTGGAGCCAGGGAAAGAA 2235  
Db 363 erSerSerAlaProAlaProProMetArgPheProTyrSerSerSerSer----- 379  
Qy 2236 GGAAGAGGTGGCGGGGAGCTGGCTCCAGCCCGCAGGATACACGAGGAAATTTAGTTGT 2295  
Db 380 -----SerSerS 382  
Qy 2296 CTCTGTCTTGTCTCAGCGTGTGAACCTCCCTCGGGCCCTTGCCTATCCAGGCCTCTCCC 2355  
Db 382 erAlaAlaAlaSer----- 386  
Qy 2356 CTTGCTTCTCCCTTCTTCCAGTTATACATCTCCCTCATCCCTTCCCTGGGCCCCAGC 2415  
Db 387 -----SerSerSerSerSerSerSerSerSerSerSerProPheProAlaSerGlnA 404  
Qy 2416 CGTCCCGCAGGGTGGAAAGGCTGTGCCCTCTCCCTATACATGCTGTCTTCCATA 2475  
Db 404 laLeuPro-----SerTyrPro-----His 411  
Qy 2476 GCCTTCTCTCTCTCTACTCATGAGACTGCCTCCATTTCTTCTTCTGCAACCTGTCTCC 2535  
Db 411 erPheProPro-----ProThrSerLeuSerValS 421  
Qy 2536 TATCAGCTGAACCTTCTTTCGAGGTGTAGTAGTACCGCTCTCTCCAGCCCTCAG 2595  
Db 421 erAsnGlnProProLysTyrThrGln-----ProSerLeuProSerGlnAlaV 437  
Qy 2596 CTGTGGGCTGGGTGTGTGTAGCGGCAATGGGGCTCTGGTTCCATGGGCACTCTCAT 2655  
Db 437 al-----TrpSerGlnGlyProProPro----- 445  
Qy 2656 CTCTCTCTTGTCTGTGTCAGAAAACCTTTGCTTCTCACTGCTCCCTCTAGTCTCC 2715  
Db 446 -----ProProTyrGlyArgLeuLeuAlaAsnSerAlaH 458  
Qy 2716 GACCTTTTCTCTCTGGCTTCTCCCTGCAAAATTTCTCAAGGAGGTGTCTACACCTC 2775  
Db 458 isPro-----GlyPro--PhePro--Pro-----SerThrGlyAl 468  
Qy 2776 TGCCTCCACTTCTCTCCACCC-----ACTCACTTCTTAACCCCTGCAATCTGGCTTC 2829  
Db 468 aGlnSerThrAlaHisProProValSerThrHisHisHisHisGlnGlnGlnG1 488  
Qy 2830 CAGGCCCCAGCAATGTTCTCTCAAGGTCTCAGGACCTCTTCCCAAGCCGACAGT 2889  
Db 488 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 506  
Qy 2890 GTTTTGAAGGCTCATTTCTCTTGTCTGTCTGTTTGCAGCCACACTGCTGAGCGCTGCTC 2949  
Db 507 -----ProGlyAl 509  
Qy 2950 CTTCTCGAATCTCTCTT-----CCTTGGTCTCTGCACT 2982  
Db 509 aPheProHis--ProLeuGlnGlyGlySerSerHisHisHisHisHisHisHisHisHis 528  
Qy 2983 CTCCTGGGCCACTTCTACCTCTCCAGCTCTCCAGGCTCTCTCTCTCTCTCTCTCTCTCT 3042  
Db 528 erProSerLeuGlySerLeuArgProTyrProProGly--ProAlaHisLeuProProp 547  
Qy 3043 CCCACAGGGGCACTCT-----CCCAAGGTTTGGCCCAACCCAGCAATCAG 3087  
Db 547 roHisSerGlnValSerTyrSerGlnAlaGlyProAsnGlyProProValSerSerS 567  
Qy 3088 CAGCTCTCTCTGAGCGTCT-----TGTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3125

Db	567	erAsnSerSerSerThrSerGlnGlySerTyrProCysSerHisProSerProSerG	587
Qy	3126	-----CCTTTTTCACGCTCTCCATTGAGAGAGTCCACCCAGCCGCACT	3168
Db	587	lnGlyProGlnGlyAlaProTyrProPhePro-ProValProThrValThrSerSer	606
Qy	3169	GCTTCAACTGTCACTGCATACAAATGATATCTTATTTGAAAAAATCAGGAGGCCATG	3228
Db	607	AlaThrLeuSerThrValIleAlaThrValAla-	617
Qy	3229	AACAAAGAACCTAGCATGGAGACAGGCGCAGTGTGAGGGGACACAAATAAGAACTT	3288
Db	618	-----SerSerProAlaGlyTyrIys-	624
Qy	3289	TGGGACAGGTATCTCTTGGTGGTGAAGCGGTCTGCCCTCTCTCTCTCCCATCAC	3348
Db	624	-----	624
Qy	3349	CCTCTCTTTTTCACAGCTGAACCTACCTGGG-CCGCCCTTTTCAACGAGCCAGACTTCAA	3405
Db	625	-----ThrAlaSerProProGlyProProProTyrGlyIysArgAlaProSer	640
Qy	3406	CCC-----TCCCGCCTGGGGGAGAGACTCTGCCAGGGCCACTGT	3447
Db	641	ProGlyAlaTyrLysThrAlaThrProProGlyTyrLysProGly-	655
Qy	3448	TGACITGGAGGTGGCGAAGCTCAATGACAACTCGCGCTGACCCAGCAACTACGAGGC	3507
Db	656	-----SerProSerPheArgThrGlyThrProProGlyTyrArgGly	670
Qy	3508	CTACAGCCACTTCTGTGTACTTGGCTGGCTTCAACCGTCAGGCTGCCACTCTGAGCT	3567
Db	671	ThrSerProPro-----AlaGlyPro-----GlyThrPheLysProGly	683
Qy	3568	GGCGCGCAG-----CCTGGCCCACTTTCGACCAAGCTTCAGGGCTGCTGGCGAGCAT	3621
Db	684	SerProThrValGlyProGlyProLeu-----ProProAla-----	695
Qy	3622	TGCGGGCGTCAATGGAGAGCTCTGGGCTACCACTGCCCCAGCCGCTGCTGGGACTGAACC	3681
Db	696	-----GlyProSerGlyLeuProSerLeuPro-----	704
Qy	3682	CACTTGGACTCTGGCCCTGGCCACAGTACTTCTCCAGAGATGGACGACTTCTGGCT	3741
Db	705	-----ProProProAlaAlaProAlaSerGlyProProLeuSer	717
Qy	3742	GCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAA	3801
Db	718	AlaThrGlnIleLysGln-----GluProAlaGlnGlu	728
Qy	3802	GAAGATGACGCTTCAGCAGCTGCAGTCACTGACCTGCACCTGGGGCTCATGGCTTCTGACT	3861
Db	729	TyrGlnThrProGluSer-----ProValPro-----	737
Qy	3862	TCTGACTTCTCTCTCTGCTCCCTTCAAACCTGTCTCCCACTTGTGAGAGCCAGCC	3921
Db	738	-----ProAlaArgSerProSerPro-----ProProLys	747
Qy	3922	CTGTATGCCAACACCT-----GTTGAGCC	3945
Db	748	Val-ValAspValProSerHisAlaSerGlnSerAlaArgPheAsnLysHisLeuAspArg	767
Qy	3946	AGGACAGACAGACTGTGAGCTCTGGCCCTTTCCTGGACCGGCTGGCGGTGATCGCAT	4005
Db	767	gclYpHeAsnSerCysAlaArgSerAspLeuTyrPheVal-----	780
Qy	4006	CAGCCCTGTCTCTCTCCCACTT-----CCCAAGGCTTACCGAGCTGG-----GG	4050
Db	791	-----ProLeuGlnGlyProSerTyrProArgSerGlyProThrTyrTyrArgCysGln	799
Qy	4051	AGGAGGTACAGTAGGCCCTGTCTGTCTGTCTTCTTACAGAAAGTCAATGCTCGAGGAGTG	4110

```

Qy 5008 CCCCTTGATCCCTCAGTCCTCCACACAGCTTCTGTC-----CACCCCAAGGTTTCCCTC 5061
      |||||:::|||
Db 1095 nProLeuLeuProHisProLeuHisGluAsnGluValLeuArgHisGlnLeuPheAlaAl 1115
      |||||:::|||
Qy 5062 ACCCCAC 5068
      :|||:::
Db 1115 aProTyr 1117
      :|||:::

```

Search completed: January 27, 2003, 16:27:02  
 Job time : 367.265 secs













Db 2006 aserGlutHislys-----ValSerPro-AlaProArgProSerSerLeuProIle----- 2022  
Qy 4915 AGCTGCACAGAGCTCTGGGAGAGGCGCTCTT-----C 4947  
Db 2023 -----LeuProSerGlyProLeuTyProGlyLeuPheAspIleArgGlyS 2038  
Qy 4948 CTCGCCGACCTGGGCGATGGCGCACCTCAGACTTACCCTGCTGCTGCCACACCAA 5007  
Db 2038 erProThrGlyGlyAlaGlySerAlaAspProPheAlaProValPheValProProH 2058  
Qy 5008 CCCCT-----TGATCCCTCCAGTC 5025  
Db 2058 isProGlyLeSerGlyGlyLeuGlyAlaLeuSerGlyAlaSerArgSerLeuSer- 2077  
Qy 5026 CTCGCACACAGCTTCTGTCACCCCA 5051  
Db 2078 --ProThrArgLeuLeuSerLeuPro 2085  
RESULT 2  
SHK1\_RAT  
ID SHK1\_RAT STANDARD; PRT; 2167 AA.  
AC Q9WV48; Q9WU13; Q9WUE8; Q9QZ8;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP  
DE interacting protein) (SPANK-1) (Synanon) (Somatostatin receptor  
DE interacting protein) (SSTR interacting protein) (SSTRIP).  
GN SHANK1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND  
RP DLG4.  
RP TISSUE=Brain;  
RC MEDLINE=99419021; PubMed=10488079;  
RA Yao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;  
RT "Synanon, a novel neuronal protein interacting with synapse-associated  
RT protein 90/postsynaptic density-95-associated protein.";  
RL J. Biol. Chem. 274:27463-27466(1999).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.  
RC STRAIN=Sprague-Dawley;  
RX MEDLINE=99360650; PubMed=10433268;  
RA Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtschanoff J.,  
RA Weinberg R.J., Worley P.F., Sheng M.;  
RT "Shank, a novel family of postsynaptic density proteins that binds to  
RT the NMDA receptor/PSD-95/GKAP complex and cortactin.";  
RL Neuron 23:569-582(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Brain;  
RX MEDLINE=20549637; PubMed=10958799;  
RA Tobaben S., Suedhof T.C., Stahl B.;  
RT "The G protein-coupled receptor Ctl1 interacts directly with proteins  
RT of the Shank family.";  
RL J. Biol. Chem. 275:36204-36210(2000).  
RN [4]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND  
RC DEVELOPMENTAL STAGE.  
RC TISSUE=Brain;  
RX MEDLINE=99436166; PubMed=10506216;  
RA Lim S., Naisbitt S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;  
RT "Characterization of the Shank family of synaptic proteins. Multiple  
RT genes, alternative splicing, and differential expression in brain and  
RT development.";  
RL J. Biol. Chem. 274:29510-29518(1999).  
RN [5]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).  
RC TISSUE=Brain;  
RX MEDLINE=20020275; PubMed=10551867;

RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;  
RT "Somatostatin receptor interacting protein defines a novel family of  
RT multidomain proteins present in human and rodent brain.";  
RL J. Biol. Chem. 274:32997-33001(1999).  
RN [6]  
RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.  
RX PubMed=10433269;  
RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,  
RA Doan A., Aakalu V.K., Lanahan A.A., Sheng M., Worley P.F.;  
RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of  
RT postsynaptic density proteins.";  
RL Neuron 23:583-592(1999).  
RN [7]  
RP INTERACTION WITH SPTAN1.  
RX PubMed=11509555;  
RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,  
RA Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;  
RT "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the  
RT multidomain Shank protein family interact with the cytoskeletal  
RT protein alpha-fodrin.";  
RL J. Biol. Chem. 276:40104-40112(2001).  
RN [8]  
RP FUNCTION.  
RX PubMed=11498055;  
RA Sala C., Piech V., Wilson N.R., Passafaro M., Liu G., Sheng M.;  
RT "Regulation of dendritic spine morphology and synaptic function by  
RT Shank and Homer.";  
RL Neuron 31:115-130(2001).  
RN [9]  
RP REVIEW.  
RX PubMed=10806096;  
RA Sheng M., Kim E.;  
RT "The Shank family of scaffold proteins.";  
RL J. Cell Sci. 113:1851-1856(2000).  
RN [10]  
RP FUNCTION: Seems to be an adapter protein in the postsynaptic  
RP density (PSD) of excitatory synapses that interconnects receptors  
RP of the postsynaptic membrane including NMDA-type and metabotropic  
RP glutamate receptors, and the actin-based cytoskeleton. May play a  
RP role in the structural and functional organization of the  
RP dendritic spine and synaptic junction. Overexpression promotes  
RP maturation of dendritic spines and the enlargement of spine heads  
RP via its ability to recruit Homer to postsynaptic sites, and  
RP enhances presynaptic function.  
RN [11]  
RP SUBUNIT: May homomultimerize via its SAM domain. Interacts with  
RP SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with  
RP DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via  
RP the PDZ domain (By similarity).  
RN [12]  
RP SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of  
RP neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.  
RN [13]  
RP ALTERNATIVE PRODUCTS: At least 5 isoforms; 1, (shown here); 2, 3,  
RP 4/a and 5; are produced by alternative splicing.  
RN [14]  
RP TISSUE SPECIFICITY: Expressed only in brain (neuropil of cortex,  
RP CAL region hippocampus and molecular layer of cerebellum).  
RN [15]  
RP DEVELOPMENTAL STAGE: Expression increases from low levels at birth  
RP to high levels at 3-4 weeks before dropping slightly in adulthood.  
RN [16]  
RP Expressed in the cortex and the molecular layer of the cerebellum  
RP at postnatal day 7. Isoform 2 expression does not change during  
RP development of both cortex and cerebellum. Isoform 4 expression  
RP decreases significantly during development of cortex but not  
RP cerebellum.  
RN [17]  
RP SIMILARITY: BELONGS TO THE SHANK FAMILY.  
RN [18]  
RP SIMILARITY: CONTAINS 7 ANK REPEATS.  
RN [19]  
RP SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.  
RN [20]  
RP SIMILARITY: CONTAINS 1 SAM DOMAIN.  
RN [21]  
RP SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
RN [22]  
RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -  
RP the European Bioinformatics Institute. There are no restrictions on its  
RP use by non-profit institutions as long as its content is in no way  
RP modified and this statement is not removed. Usage by and for commercial  
RP entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
RP or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```

CC DR EMBL; AF102855; AAD04569.2; --
DR EMBL; AF131951; AAD29417.1; ALT_INIT.
DR EMBL; AF159045; AAD42975.1; --
DR EMBL; AF141904; AAF02498.1; ALT_INIT.
DR HSSP; PF0519; IABL.
DR DR InterPro; IPR002110; ANK.
DR DR InterPro; IPR001478; PDZ.
DR DR InterPro; IPR001660; SAM.
DR DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; ank; 7.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00536; SAM; 1.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS0088; ANK_REPEAT; 1.
DR PROSITE; PS0297; ANK_REPEAT; 3.
DR PROSITE; PS0002; SH3; 1.
DR PROSITE; PS0105; SAM_DOMAIN; 1.
DR ANK repeat; SH3 domain; Repeat; Alternative splicing.
FT REPEAT 195 210 ANK 1.
FT REPEAT 212 245 ANK 2.
FT REPEAT 246 278 ANK 3.
FT REPEAT 279 312 ANK 4.
FT REPEAT 313 345 ANK 5.
FT REPEAT 346 378 ANK 6.
FT REPEAT 379 395 ANK 7.
FT DOMAIN 554 613 SH3.
FT DOMAIN 663 757 PDZ.
FT DOMAIN 2104 2167 SAM.
FT DOMAIN 929 932 POLY-PRO.
FT DOMAIN 1010 1015 POLY-HIS.
FT DOMAIN 1022 1027 POLY-HIS.
FT DOMAIN 1194 1199 POLY-GLY.
FT DOMAIN 1850 1860 POLY-PRO.
FT VARSPPLIC 1 613 MISSING (IN ISOFORM 2).
FT VARSPPLIC 615 654 SOEGROESDKAKRLFRHYTVYDSFAPSLIDGDSG
FT VARSPPLIC 646 654 MISSING (IN ISOFORM 2).
FT VARSPPLIC 797 804 MISSING (IN ISOFORM 3).
FT VARSPPLIC 1930 1943 MISSING (IN ISOFORM 4).
FT VARSPPLIC 1944 2167 LSEDSTLSKPS -> QYRIIVKSSDFGDF (IN
FT VARSPPLIC 1141 1141 MISSING (IN ISOFORM 5).
FT CONFLICT 1174 1174 S -> T (IN REF. 1).
FT CONFLICT 1246 1246 S -> N (IN REF. 2).
FT CONFLICT 1246 1246 R -> K (IN REF. 1).
FT CONFLICT 1323 1323 A -> T (IN REF. 1).
FT CONFLICT 1331 1331 S -> D (IN REF. 1).
FT CONFLICT 1726 1726 S -> N (IN REF. 2).
SQ SEQUENCE 2167 AA; 226333 MW; 3F478B5A7B18BA86 CRC64;

Alignment Scores:
Pred. No.: 4,98e-09 Length: 2167
Score: 351.00 Matches: 334
Percent Similarity: 29.90% Conservative: 102
Best Local Similarity: 22.91% Mismatches: 483
Query Match: 3.72% Indels: 541
DB: 1 Gaps: 69

US-09-931-704-3 (1-5087) x SHK1_RAT (1-2167)
QY 1084 CACATGCCAAAGACTCTC-----TCAGCTGACACACAGATCCATTCACAA 1128
Db |||||
Db 858 HisArgProLysGlyPhePheAlaThrGluSerSerPheAspProHisHisArgSerGln 877
QY 1129 GTATCTACTAGACACTCATCGTGCCA-----|||||

```

```

Db 878 ProSerTyrAspArgProSerPheLeuProProGlyProGlyLeuMetLeuArgGlnLys 897
QY 1159 -----AGTCCTCATCTCTCAAAACATACACA 1182
Db |||||
Db 898 SerTleGlyAlaAlaGluAspArgProTyrLeuAlaProProAlaMetLysPheSer 917
QY 1183 TGCCTCTCTTTCTCTCCCGTCTTCCAGAGTGTTCCTCCCTCCATCC-----C 1233
Db |||||
Db 918 ArgSerLeuSerValProGlySerGluAsp-IleProProProThrThrSerProP 937
QY 1234 CTCTGCTCTCCATCTGGTGTCTCCACCTCACCCACCCAGTCCAGTGGGGACAGAC 1293
Db |||||
Db 937 roGluProProTyr-SerThrProAlaProSer-SerGlyArgLeuThrProSe 956
QY 1294 ACCTGAGGGGCTGCAGCTGCTCCCGTGTGGGGCCGGCGCTCATGCTTCTCGTC 1353
Db |||||
Db 956 rProArgGlyGlyPro---PheAsnProSerSerGlyGlyPro-LeuProAlaSerSerP 975
QY 1354 CATCC-----TGCCACAGGGGACTCTGTGGGGATGTTAGCTGCTCT--- 1395
Db |||||
Db 975 roSerSerPheAspGlyProSerProProAspThrArgGlyGly-GlyArgGlyLysSer 994
QY 1396 ---GTGCACGGTGTCTGGCACCTCCCTGC-----AGTGCAGCTCTCAATCGCAC 1443
Db |||||
Db 995 LeuTyrHisSerAlaAlaLeuProProAlaHisHisHisProHisHisHisHis 1014
QY 1444 AGGGGACCCAGGGCTGGCCCTCCATCCAGAAAACCTATGACCTCACCCG----- 1494
Db |||||
Db 1015 HisAlaProProProGlnProHisHisHisAlaHisProProHisProGluMet 1034
QY 1495 -----CTACCTGGAGCACCACTCCGACGCTGG-CTGGGACCTATGTGAGTA 1541
Db |||||
Db 1035 GluThrGlyGlySerProAspAspProProProArgLeuAlaLeuGlyProGln----- 1052
QY 1542 TCCAGCTAGGAATCTGGAGTGTGGGAGGAGTAGGAGTGGGGAAAGACAGCTCTAAC 1601
Db |||||
Db 1053 ProSerLeuArgGlyTrp----- 1058
QY 1602 CGTGGAGGGTTCTGTAAATGATGGGTGAGGAGGGCTCTTTGGCTCCACACAGTCC 1661
Db |||||
Db 1059 ArgGlyGly----- 1065
QY 1662 CTGTCTGGTCTATCTCTGCTCCCTTCCCTTCTAGTGGCCCCCCCCCTCTCCCATCCTCTGG 1721
Db |||||
Db 1066 ThrSerGly-----AlaProSerProSerHisSer 1076
QY 1722 CCCAGAGTAGGATGTGGGCGCCCTCGCACCCGCT-----TGGCCCATGTC 1772
Db |||||
Db 1077 SerSerGlyGlySerSerGlyProThrGlnAlaProAlaLeuArgTyrPheGlnLeuPro 1096
QY 1773 CCAGTGGTGGCAGCCCGCCGCTCCCTGGGGCGCGGAGTCTCTCTCTGT 1832
Db |||||
Db 1097 ProArgAlaAlaSerAlaAlaMetTyrValProAlaArgSerGlyArgGlyGly 1116
QY 1833 TACACCGTGTGTGTGTCTCTTGGCGGGGGGTGTGGTGGGGACA----- 1880
Db |||||
Db 1117 ProLeuValLysGlnThrLysValGluGlyGluProGlnLysGlySerIleProSerAla 1136
QY 1881 GAGGGGCCCCACCTCCCATGCTGGTTCAGCTCGCTCTGCCCCACACCTGGGGCCC 1940
Db |||||
Db 1137 SerSerProThrSerProAlaLeuProArgSerGlu-----ProProProAlaGlyPro 1154
QY 1941 TGCTGCTCTGGACCCAGGGGCTCCCTTCCGCTCTCCCTCTCCCATCTAGCTGGGGCTCC 2000
Db |||||
Db 1155 SerGluLysAsnSerIleProIleProThrIleIleLysAlaProSerThrSerSer 1174
QY 2001 TAGGGGGTCTATGGGGGAGGGGACTGTAGGGA-----ACC 2036
Db |||||
Db 1175 -----SerGlyArgSerSerGlnGlySerSerThrGluAlaGluProProThr 1190
QY 2037 CAGGCAGTAGTGGCAGGGGGTTAGGGTGTGGATGGAGGTATGCTGTAAAGATTGGGG 2096
Db |||||
Db 1191 GlnProAspGlyAlaGlyGlyGly----- 1199

```



```
Db 1855 robProLeuProGlyProLeuSerGlnProGlnAlaSerAlaLeuAlaThrValLysAla 1875
QY 3603 ----- ||||| |||||
Db 1875 erlleleSerGluLeuSerSerLysLeuGlnGlnPheGlyGlySerThrAlaGlyG 1895
QY 3604 -----GGGGCTGTCTGGCGCAGCATTCGGGGCTCATGGC----- 3636
Db 1895 lyAlaLeuProTtpAlaArgGlyGlySerGlyGlySerThrAspSerHisGlyGlyA 1915
QY 3637 ----- ||||| |||||
Db 1915 laSerTyrilleProGluArgThrSerSerLeuGlnArgGlnArgLeuSerGluAspSerG 1935
QY 3638 GCCTCTGGCTTACCACTGCCCGCCGCTGCTGGAGTGAACCACTTGACCTCTGGC 3697
Db 1935 lnThrSerLeuSerLysProSerSer-----SerIlePheGlnAsnTrp 1951
QY 3698 CTTGCCACAGTACTCTCTCCAGAGATGAGCAGCTTCTGGCTGCTCAAGGAGCTGCAG 3757
Db 1951 roLysProPro-----LeuProPro-----LeuProThrGlySerGlyValSerS 1966
QY 3758 ACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAAAGATGACAGCTCCA 3817
Db 1966 erSerThrAlaAlaAlaProGlyAlaThrSerProSer-----AlaSerSerAlaSerA 1984
QY 3818 GCAGCTGCAGTCACTCCCTGCGCGGCTCATGCTTCTGACCTTCTGACCTCTCTCTCT 3877
Db 1984 laSerThrArgHis-----LeuGlnGlyValGluPheGluMetArgProProLeuLeuArgA 2003
QY 3878 TCCTCTCCC-----CTTCA-----AACCTCTCTCCCACTT 3907
Db 2003 rgAlaProSerProSerLeuLeuProAlaSerAspHisLysValSerProAlaProArg- 2022
QY 3908 TGTGAGGAGCAGCCCTGTATGCCAACA-----CCTGTTGAGCAGCAGCAGAGCTGTGAG 3964
Db 2023 ProSerSerLeuProLeuLeuProSerGlyProIleTyrProGlyLeuPheAspIleArg 2042
QY 3965 CCTCTGGCCCTTCTCTGGACCGCTGGCGCTGTGATCGATCGATCCCTGTCTCTCCCA 4024
Db 2043 SerSerProThrGlyAlaGlyGlySerThrAspProPheAlaProValPheValPro 2062
QY 4025 CTTCCCAAGAGTTCAGGAGCTGGGAGGAGGTACAGTAGCCCTGCTCTCTCTGTTTC 4084
Db 2063 ProHisProGlyIle-----SerGlyGlyLeuGlyGlyAlaLeuSer----- 2076
QY 4085 TACAGGAAGTCATGCTCGAGGAGTGTGAAGTGTGAGTGTGAGGAGGAGGCTCATG 4144
Db 2077 -----GlyAlaSerArgSerLeuSerProThrArgLeu----- 2087
QY 4145 GCCTCTGTCTTCTTGCTTACCACTTGGCCAGTGGCCACCGCCCTCAGGTGGCACATC 4204
Db 2088 -----LeuSerLeuProProAspLysProPheGlyAlaLysProLeu----- 2101
QY 4205 TGGAGGCGAGGGTGTGAGGGGCCACCAACACATGCTTCTTGGGTGAA----- 4255
Db 2102 -----GlyPheTrpThrLysPheAspVal 2109
QY 4256 GCCTTTTGGCTGCCCACTCTCTCTGATGGGTGTGCTCTCTCTATCCCAATCACTCT 4315
Db 2110 AlaAspTrpLeu-----GluTrpLeuGlyLeuSerGluHisArgAlaGlnPheLeu 2126
QY 4316 ATACATCCATCAGGAACAACATGTTGGCAATT 4351
Db 2127 AspHisGluIleAspGlySerHisLeuProAlaLeu 2138
RESULT 3
CA13 HUMAN
ID CA13 HUMAN STANDARD; PRT; 1466 AA.
AC P02461; Q15112;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
```

```
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin fibroblast;
RX MEDLINE=89350838; PubMed=2764886;
RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA Prockop D.J.;
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT chain of human type III procollagen. Differences in protein structure
RT from type I procollagen and conservation of codon preferences.";
RL Biochem. J. 260:509-516 (1989).
RN [2]
RP SEQUENCE OF 149-1225 FROM N.A.
RX MEDLINE=89386015; PubMed=2780304;
RA Janeczko R.A., Ramirez F.;
RT "Nucleotide and amino acid sequences of the entire human alpha 1
RT (III) collagen.";
RL Nucleic Acids Res. 17:6742-6742 (1989).
RN [3]
RP SEQUENCE OF 168-398.
RX MEDLINE=77134724; PubMed=5573335;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of cyanogen
RT bromide peptides from the amino-terminal segment of type III collagen
RT of human liver.";
RL Biochemistry 16:1158-1164 (1977).
RN [4]
RP REVISIONS.
RA Seyer J.M.;
RL Submitted (DEC-1977) to the PIR data bank.
RN [5]
RP SEQUENCE OF 399-727.
RX MEDLINE=79000343; PubMed=687591;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of five
RT consecutive CNBr peptides from type III collagen of human liver.";
RL Biochemistry 17:3404-3411 (1978).
RN [6]
RP SEQUENCE OF 728-964.
RX MEDLINE=80198282; PubMed=6246925;
RA Seyer J.M., Mainardi C., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha 1
RT (III)-CB5 from type III collagen of human liver.";
RL Biochemistry 19:1583-1589 (1980).
RN [7]
RP SEQUENCE OF 950-1466 FROM N.A.
RX MEDLINE=88189827; PubMed=3357782;
RA Mankoo B.S., Daigleish R.;
RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
RL Nucleic Acids Res. 16:2337-2337 (1988).
RN [8]
RP REVISION TO 1184.
RX MEDLINE=89098346; PubMed=3211760;
RA Molyneux K., Daigleish R.;
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL Nucleic Acids Res. 16:11833-11833 (1988).
RN [9]
RP SEQUENCE OF 1065-1466 FROM N.A.
RX MEDLINE=85087944; PubMed=6096827;
RA Leidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA Rosenbloom J., Myers J.C.;
RT "Molecular cloning and carboxyl-propeptide analysis of human type III
RT procollagen.";
RL Nucleic Acids Res. 12:9383-9394 (1984).
RN [10]
RP SEQUENCE OF 965-1200.
RX MEDLINE=81208139; PubMed=7016180;
```





QY 4891 ---GGAGCCAGGGTGGCAAGAGTCTGATGACACACAGGAGGGGTGAGCTGGGCAAGA 4835  
Db |||||  
260 ProGlyMetLysGlyHisArgGly---PheAspGlyArgAsnGlyGluLysGlyGlu--- 277  
QY 4834 GACTAAGTCTGGGCAAGCATTTTATTGTTAATACAGAATAGAAATTCGCAATAAA 4775  
Db |||||  
277 --- 277  
QY 4774 TATCATCTAATAATAACATCTCCAATAATAATAATAATAATAACACACACTTAGAGTCA 4715  
Db --- 277  
QY 4714 TGAGTGGTGGGCTGGGGGCGAGGGCTGGGGAGCTGCCA--- 4673  
Db |||||  
278 ---ThrGlyAlaProGlyLeuLysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaPro 296  
QY 4672 ---CCCTACACCCCAAAATCTACTGATGTAACCTTTCAGGAATCTCTGGTGTGGC 4616  
Db |||||  
297 GlyProMetGlyProArg--- 302  
QY 4615 TATGGTTGGCTCCCAAGCTCGCAACCCACAGATACCTGGGAAAGGGGCGACAGAAGA 4556  
Db |||||  
303 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg--- 320  
QY 4555 GGCACCGTAATCTGGAAGACAATTCAGGCAAGGTCTGATGCTCAGCTCGGGTTTGT 4496  
Db |||||  
321 ---GlyAsnAspGlyAlaArgGlySerAspGlyGlnProGlyProPro 335  
QY 4495 TTGCCACTGTCTTGGAGACTCTCTGGTCTGGTCTGGTCTCTCTCTCTCTCTCTCTCT 4436  
Db |||||  
336 GlyProProGlyThrAlaGlyPheProGly--- 345  
QY 4435 TTCAGTATTTCCT 4376  
Db |||||  
346 ---SerProGlyAlaLysGlyGluValGlyProGlyProAlaGlySerPro--- 359  
QY 4375 TTAATCTCATCTCTTTTGTGTAGAAATGCCACCATGTTTGTCTTCTGAAATGGATGAT 4316  
Db --- 359  
QY 4315 AGAGTGATTTGGGATAAGGAGCAACACCCATCCAGGAGGTGGGCGACCAAGGGC 4256  
Db |||||  
360 ---GlySerAsnGlyAlaProGlyGlnArgGlyGluPro--- 371  
QY 4255 TTCACCCCAAGAAAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4196  
Db |||||  
372 ---GlyProGlnGlyHisAlaGly---AlaGlnGlyProProGly--- 384  
QY 4195 CCTGAGGGCTGGTGGGCACTGGCCAACTGCTAGGCAAGAGCAGGAGCCATGAGCGC 4136  
Db |||||  
385 ProProGlyIleAsnGlySer---ProGlyGlyLysGlyGluMetGly--- 399  
QY 4135 CTCTGCACCAACCTGAACCACTTCACACTCCCTCGACATGACTTCCTGTAGAAACAGGA 4076  
Db |||||  
400 ---ProAlaGlyIleProGlyAla---ProGlyLeuMetGly 411  
QY 4075 CAGGACAGGCTCTACTGTACTCTCTCCAGCTCGGTAGACCT---TTGGGAGGTGG 4022  
Db |||||  
412 -AlaArgGlyProProGlyProAlaGlyAlaAsnGlyAlaProGlyLeuArgGlyAla 431  
QY 4021 GGAGGAGACAGGCTGATCGCATCACAGCCCGGCTCCAGAAAGGGCGCAGAGGCTC 3962  
Db |||||  
431 aGlyGluProGlyLysAsnGlyAlaLysGlyGluProGlyProArgGlyGluArgGly--- 450  
QY 3961 ACAGCTTCT 3905  
Db |||||  
451 ---GluAlaGlyIleProGlyValProGlyAlaLysGly-GluAspGlyLysA 467  
QY 3904 TGGGACACAGGTTTGAAGGGGCGGAAGAGGAGAGGTGAGAAGTCCAGAACCCATGAGC 3845  
Db |||||  
467 spGlySerProGlyGluProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyA 487  
QY 3844 CQCCAGGT-----GCAGGTGACTGCGAGCTGCTGGAGGCTGCTCTCTCTCT 3797

Db |||||  
487 laProGlyPheArgGlyProAlaGly--- 495  
QY 3796 GAGCCGGTTGAAGTCTTGGCCGAGGCGCCACAGCCAGGCTCTGAGCTCTTTCAGACCCA 3737  
Db |||||  
496 ---ProAsnGlyIleProGlyGluLysGlyProAlaGlyGluA 509  
QY 3736 GAAGTCGTCTCTTCTTGGAGGAAGTCAGTGTGGGCGAGGCGCCAGGAGTCCCAAGTGGGTT 3677  
Db |||||  
509 rGlyAlaProGlyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArg--- 526  
QY 3676 AGTCCAGGCGAGCTGGGCGAGTGGGTAGCCAGAGCTGCCATGACGCCGCCAATGCT 3617  
Db |||||  
527 ---AspGlyValProGlyGlyProGlyMetArgGlyMetProGly---S 541  
QY 3616 GCCAGAGCGCTCGAGGCTGTGAGAGTGGGCGAGGCTGCGGCGAGCTCAGCAGT 3557  
Db |||||  
541 erProGlyGlyProGlySerAspGlyLysProGlyProProGlySerGlnGlyGluSerG 561  
QY 3556 GGCAGCTGAGCTGTGAGGCGACGCAAGTAACACAGAAGTGGCTGTAGGCTCGTAGTT 3497  
Db |||||  
561 lyaArgPro---GlyProProGlyProSerGlyProArg--- 572  
QY 3496 CTGGTCAAGCCAGTCTTTCATGAGGCTTCGCCACACCTCCAACTCAACAGTGGCCCT 3437  
Db |||||  
573 -GlyGlnProGlyValMet---GlyPheProGlyPro-Lys--- 584  
QY 3436 GGGCAGAGCTCTGCCCGCGGGGAGGTTGAAGTCTGGCTCTTGAAGGGGGGCC 3377  
Db |||||  
585 -GlyAsnAspGlyAlaProGlyLysAsnGlyGluArgGlyGly---ProGlyGlyPr 602  
QY 3376 CAGGTAGTTACAGTGTGAAAAGAGAGGCTGATGGGAAGGAGGAGGCGAGCGCTGG 3317  
Db |||||  
602 o---GlyProGlnGlyProProGlyLysAsnGlyGluThrGlyProG 617  
QY 3316 CTCACCAACAGGAGATACCTGCTCCCAAGTTCTATTTTGTGTGCTCCCT--- 3265  
Db |||||  
617 nGlyProPro---GlyProThrGlyPr 625  
QY 3264 ---GACACTGGCCCTGTCTCCATGTAGTCTTCTTGTTCATGGCCT 3221  
Db |||||  
625 oGlyGlyAspLysGlyAspThrGlyProProGlyProGlnGly---LeuGlnGlyLe 643  
QY 3220 CCTCAGTCTTTTCCAAATAAGGATATCATTTGTATGACAGGTGACAGTTGAAGCAGTGGCGG 3161  
Db |||||  
643 uPro---GlyThrGlyG 648  
QY 3160 TGGTGAGCTCTCCAATGGAGAGGCGTA-GAATAAGAGAGGAGGAGAGAGCACAAGACGC 3102  
Db |||||  
648 yProProGlyGluAsnGlyLysProGlyGluProGlyProLysGlyAspAla--- 665  
QY 3101 TCAGGAAGGACGTGTGATGCTGGTGGGCAACCTTGGGAGAGTGGCCGCTGTGGGG 3042  
Db --- 665  
QY 3041 GCAGGACAGAGGAGAGGAGGCTGGAGAGGTAGAGGTGGCCGAGGAG 2982  
Db |||||  
666 ---GlyAlaProGlyAlaProGlyGlyLysGlyAspAla---GlyAlaProGlyG 682  
QY 2981 GTGAGAGACCAAGAGAGGAGTTTCGAAAGGAGGAGGAGCGCTCAGCAGTGTGGCTGCAA 2922  
Db |||||  
682 uArgGlyProProGlyLeuAlaGlyAlaProGly--- 693  
QY 2921 AACAGACAGCAAGGAGATGAGCTTCAAAACACTGTCTGGGCTTGGCAAGGAGTCCCTG 2862  
Db |||||  
694 ---LeuArgGlyGlyAlaG 699  
QY 2861 ACAGCTTGGAGAGAACCATCTCTGGGCTTGGAGCCAGATTCAGGGGGTTAAGAAGT 2802  
Db |||||  
699 yProPro---GlyPro---GluGlyGlyLysGlyAl 709  
QY 2801 GAGTGGGTGGAGAGGAGTGGAGGAGGAGGTTGTAGACCACTCTCTTGGAGAAATTTGGCA 2742  
Db |||||



"Identification of the promoter and first exon of the mouse alpha 1 (III) collagen gene.";  
 J. Biol. Chem. 260:3773-3777(1985).  
 [4]  
 SEQUENCE OF 810-1464 FROM N.A.  
 STRAIN=C57BL/6J; TISSUE=Embryonic head;  
 MEDLINE=21085660; PubMed=11217851;  
 Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehi Y., Arakawa T., Hara A., Fushimi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I., Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R., Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Baran G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.;  
 "Functional annotation of a full-length mouse cDNA collection.";  
 Nature 409:685-690(2001).  
 [5]  
 SEQUENCE OF 1442-1464 FROM N.A.  
 STRAIN=C57BL/6;  
 MEDLINE=91274355; PubMed=2054384;  
 Metaseranta M., Toman D., de Crombrughe B., Vuorio E.;  
 "Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.";  
 Biochim. Biophys. Acta 1089:241-243(1991).  
 CC -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 CC ALONG WITH TYPE I COLLAGEN.  
 CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE  
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE  
 CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC O-LINKED GLYCANS CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 1 VFNC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X52046; CAA36279.1; -;  
 DR EMBL; M18933; AAA37338.1; -;  
 DR EMBL; K03037; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AK019448; BAB31724.1; -;  
 DR EMBL; X57983; CAA41048.1; -;  
 DR PIR; A22287; A22287.  
 DR PIR; A27353; A27353.  
 DR PIR; S16373; S16373.  
 DR MGD; MGI:88453; Col3a1.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000885; Fib collagen\_C.  
 DR InterPro; IPR001007; VFNC.  
 DR Pfam; PF01391; Collagen\_18.  
 DR Pfam; PF01410; COLFI.1.  
 DR ProDom; PD000007; Collagen; 1.  
 DR ProDom; PD002078; Fib collagen\_C; 1.  
 DR SMART; SM00208; COLFI; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS01208; VFNC; 1.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW

KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.  
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.  
 FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.  
 FT DOMAIN 31 90 VFNC.  
 FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).  
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION (C-TERMINAL).  
 FT DOMAIN 1196 1464 NONHELICAL REGION (C-TERMINAL).  
 FT CARBOHYD 262 262 O-LINKED (GAL. .) (BY SIMILARITY).  
 FT MOD RES 262 262 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 283 283 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 859 859 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 976 976 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 1093 1093 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 1105 1105 HYDROXYLATION (BY SIMILARITY).  
 FT DISULFID 1195 1195 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 1196 1196 INTERCHAIN (BY SIMILARITY).  
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;  
 Alignment Scores:  
 Pred. No.: 7.29e-08 Length: 1464  
 Score: 325.00 Matches: 377  
 Percent Similarity: 28.83% Conservative: 73  
 Best Local Similarity: 24.15% Mismatches: 549  
 Query Match: 3.48% Indels: 562  
 DB: 1 Gaps: 83  
 US-09-931-704-3 (1-5087) x CA13\_MOUSE (1-1464)  
 QY 4672 CCTACACCCCAAAATGCTACTGCTTAACTTTTCAGGAATCTGTGGTGGCTAT 4613  
 DB 92 ProSerThrPro-----AlaProValLeuProAspGlyHis 103  
 QY 4612 GGTTCCTCCCTCCAG-----CTGGCAACCCACAGATACCTCGGGAAGG 4568  
 DB 104 Gly-----ProGlnGlyProLysGlyAspProGlyProGlyIleProGlyAsn 121  
 QY 4567 GGGCAGAGAGAGGACCGCTAATACTGGAAGACATTCGAGGCAAGTCTGTGCTCAG 4508  
 DB 122 GlyAsp-----ProGlyLeuProGlyGln 129  
 QY 4507 CTCGGGTTTGTTCCTCCACTCTGTCTGTGGAGACTCTCGGTCTGTCTGTCTGT--- 4451  
 DB 130 ProGlyLeuProGlyPro-----ProGlySerProGlyIleCysGlu 143  
 QY 4450 CCTTCCTCCCTGTGCTTTCAGTATTTCTCTTCGTGTTATAGGCACTCCCAATGAGAC 4391  
 DB 144 SerCysProThr-----GlyGlyGlnAsnTyrSer 153  
 QY 4390 CCCAACCTCGCACTGTAACTCTATCTCTTTTGTGTAGAAATGCCACCATGTTGTTTC 4331  
 DB 154 ProGln-----PheAspSer----- 158  
 QY 4330 CTGAATTGGATGTATAGAGTATTTGGGATAAGGAGCAACACCATCCAGGAGAGTG 4271  
 DB 159 -----TyrAspValLysSerGly-----Val 165  
 QY 4270 GGGCAGCCAAAGGGCTTCACCCCAAGAGGCAATGTGTGTGGTGGCCCTCAACCCCTGC 4211  
 DB 166 GlyGlyMetGlyGlyTyrPro-----GlyProAlaGlyProPro 178  
 QY 4210 CCTCCAGATGTGCCACCTGAGGGGCTGGTGGGCACTGGCCAAAGTGGTAGGCAAGAAGCA 4151  
 DB 179 GlyProProGlyProPro---GlySerSerGlyHis-----ProGlySerProGlySer 195  
 QY 4150 GGAGGCCATGAGCGCTCTCGACCAACCACTTCCACTTCCACTCCCTCGAGCATGACTT 4091  
 DB 196 ProGlyTyrGlnGly-----Pro 201  
 QY 4090 CCTGTAGAAACAGGACAGGAGCGGCTACTGTACTCTCTCCCTCCAGCTGGTAGACTTT 4031  
 DB 202 ProGlyGluProGlyGln-AlaGlyProAlaGlyProProGlyPro---AlaLe 220

QY 4030 GGGAGGTGGGAGGAGCAGGGCTGATCGCATCACAGCCCGCCGCTCCAGAAAGGGC 3971  
DB 220 uGlyProAlaGlyProAlaGlyLysAspGlyGluSerGlyArgProGlyGlu 240  
QY 3970 CAGAGGTACAGCTTCTGCTCTGCTGCTCAACAGGTGTT-----GGCATACA 3923  
DB 240 uArgGly-----LeuProGlyProProGlyLysGlyProAlaGlyMetPr 256  
QY 3922 GGGCTGGCTCTCACAAAGTGGGAGCAGGTTTGAAGGGGAGCGAAGAGGAGGTGAC 3863  
DB 256 oGlyPheProGlyMetLysGlyHisArgGlyPheAspGlyArgAsnGlyGluLys-GlyG 276  
QY 3862 AAGTCAGAAAGCCATCAGCCCGCCAGGTGCGAGGTGCTGCTGCTGCTGCTGCTGCTGCT 3803  
DB 276 luThrGlyAla-----ProGlyLeu-----LysGlyGluA 286  
QY 3802 CTTCTGAGCCGCTTGAAGTCTTGGCGGAGCGCCACAGCAGGTCTGCTGCTGCTGCTGCT 3743  
DB 286 snGlyLeuProGlyAspAsnGlyAlaProGlyProMetGlyProArgGlyAlaPro---G 305  
QY 3742 CAGCCAGAGTCTGCTCTCTGAGGAAGTCACTGTGGCGAGGCGCCAGGAGTCCAGT 3683  
DB 305 lyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArgGlyAsnAsp----- 322  
QY 3682 GGGTTCAGTCCAGCGCGCTGGGCGAGTGGGTAGCCCGAGAGCTGCCATGACGCCCGC 3623  
DB 323 -----GlyAlaArgGlySerAspGlyGlnProGlyProProGlyProProG 338  
QY 3622 AATGCTGC-----CCAGCAGCCCTGAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 3579  
DB 338 lyThrAlaGlyPheProGlySerProGlyAlaLysGlyGluValGlyProAlaGlySerp 358  
QY 3578 -----GGTGGCGGCAGCTCAGCAGTGGCAGCTGCGGTGAGCGGTGAGCGCAAGTGA 3527  
DB 358 roGlySerAsnGlySerProGlyGlnArgGlyGluProGlyProGlyGlnGlyHisAlaGly 378  
QY 3526 ACACAGAGGTGGTGTAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3489  
DB 378 laGlnGlyProGlyProGlyProGlyAsnAsnGlySerProGlyGlyLysGlyGluMetG 398  
QY 3488 -----GCCGAGTTTGTCAATTGAGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 3446  
DB 398 lyProAlaGlyProGlyAlaProGlyLeuLeuGlyAlaArgGlyProProGly----- 416  
QY 3445 AGTGGCTGGCGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3386  
DB 417 -----ProAlaGlyThrAsnGlyLysProGlyThrArgGlyProSer----- 430  
QY 3385 AGGGGGCCCGAGTGTAGTTCAGCTGTGAAAGGAGAG-----GGTGAT---GGG 3341  
DB 431 --GlyGluProGlyLysAsnGlyAla-LysGlyGluProGlyAlaArgGlyGluArgGly 449  
QY 3340 GAAGGAGGAGGAGCGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3281  
DB 450 GluAlaGlySerProGly-----LysGlyProLys----- 460  
QY 3280 ATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3221  
DB 460 ----- 460  
QY 3220 CCTGAGTTTTTCCAAAGGATATCAITTTGTATGTCAGGTGACAGTTGAAGCAGTGGCGG 3161  
DB 460 ----- 460  
QY 3160 TGGTGAGCTCTCAATGAGAGGCGCTAGAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 3101  
DB 461 -----GlyGluAspGlyLysAspGly 467  
QY 3100 CAGGAGAGCGT-----GCTGATTGGTGGTGGGCAACCTTGGGAGAGTGGCGGCTG 3047  
DB 468 SerProGlyGluProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyProSer 487

QY 3046 TGG-----GGCAGGAGCAGAGAG 3029  
DB 488 GlyPheArgGlyProAlaGlyProAsnGlyLysGlyGluLysGlyProGlyGlu 507  
QY 3028 GAAGAGAGCTGAGAGAGCTGGAGAG-----CTAGAAAGTGGCCAGAGAGTGCAGA 2975  
DB 508 ArgGlyGlyProGlyProAlaGlyProArgGlyValAlaGlyGluProGly-----Arg 525  
QY 2974 GAC-----CAAGGAAGAGGAGTTCGAGAGGAGCAGCGCTCAGCAGTGTGGCTGC 2924  
DB 526 AspGlyThrProGlyGlyProGlyLysLeuArg---GlyMetProGlySerProGlyGlyPro 544  
QY 2923 AAAACAGCAGCAGCAGAGAGAGTACGCTTCAAAACACTGTCTGGCTGGCAGAGAGTGC 2864  
DB 545 GlyAsnAspGlyLysProGlyProGlySer-----GlnGlyGluSer 559  
QY 2863 TGACGACTTGGAGAGAACCATTCCTGGGCTGAGAGCCAGATTCAGGGGTTAAGAA 2804  
DB 560 GlyArgProGlyProProGlyProSerGlyProArgGlyGlnProGly----- 575  
QY 2803 GTGAGTGGGTGGAGAGGAGTGGAGGAGGAGGTGTAGCACCCTCTTGGAGAAATTTGG 2744  
DB 576 ValMetGlyPheProGlyProGlyLysGlyAsnAspGlyAlaPro----- 589  
QY 2743 CAGGGAAGCCAGGAGAGAA-----AAGGT 2717  
DB 590 --GlyLysAsnGlyGluArgGlyGlyProGlyGlyProGlyLeuProGlyProAlaGly 608  
QY 2716 CGGGAACAGAGAGGAGGAGTGGAGTGAAGCAAGATTTCTTGCACAGGAGACAGAGAGA 2657  
DB 609 LysAsnGlyGluThrGlyProGlyProGlyProGlyProGlyProGlyProGlyProGly 628  
QY 2656 GATCAGAGTGGCCATTTGGAAACACAGAG-----CCCCATTTGGCCCTCAGCACCAGC 2603  
DB 629 GlyAspGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGly 647  
QY 2602 CCACGAGTGGGGCTGGGAGAGAGCGGTACTACTAACTCCGAAAGAGGGTTCA 2543  
DB 648 ProProGlyGluAsn--GlyLysProGly----- 656  
QY 2542 GCTCATAGGAGCAGGTTGACAGAGGAGAAATGAGGAGCTCTCATGATGAGACAGA 2483  
DB 657 -----GluProGlyProLysGly-GluValGlyAlaProGlyAlaPr 670  
QY 2482 GGAAGCTTATGGAAGACAGCATGTTATAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2423  
DB 670 oGlyGlyLysGlyAspSerGlyAlaProGlyGluArgGlyProPro----- 685  
QY 2422 GGGAGCGCTGGGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2363  
DB 686 -GlyThrAlaGlyLysProGlyAlaArgGlyGlyAlaGlyProProGlyProGlyProGly 704  
QY 2362 AAGCAAGGAGGAGCGCTGGGATAGGCAAGGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAG 2303  
DB 705 -----GlyLysGly----- 707  
QY 2302 CACAGAGCAAACTAAATTTCTCGTGTATCTCTGGGCTGGAGCCAGCT-----CCCC 2249  
DB 708 -----ProAlaGlyProProGly 713  
QY 2248 GCCACCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2189  
DB 713 yProGlyAlaSerGlySerProGlyLeuGlnGlyMet-----ProGlyGluAr 730  
QY 2188 ACAGGGAAGAGGCGCGCTTCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2143  
DB 730 sGlyGlyProGlySerProGlyProLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 750  
QY 2142 -ACCT 2090  
DB 750 aAspGlyValProGlyLysAspGlyProArgGlyProAlaGlyProAlaGlyProProGly 770  
QY 2089 TCCTTACAGCATAACT 2032

Db	770	yProAlaGlyGlnProGlyAspLysGlyGluGlyGlySerProGlyLeuProGlyIleAl	730
Qy	2031	-----CCTFACAGTCCCTTCCTCCCATGACCCCTCCTAGGAGGCCAGCTA	1988
Db	790	aGlyProArgGlyGlyProGlyGluArgGlyGluHisGlyProGlyProAlaGlyPh	810
Qy	1987	GGATGGGAGGCGACAGCGAAGAGGGAGCCCTGGGTCCAGACAGCAGCAGG---	1931
Db	810	eProGlyAlaProGlyGlnAsnGlyGluProGlyAlaLysGlyGluArgGlyAlaProG	830
Qy	1930	TCTGGGGCAGAGGCGAGC-----TGGAAACGAGCG-----ATGGAGGTGGGCCCTCT	1880
Db	830	yGluLysGlyGluGlyGlyProProGlyProAlaGlyProThrGlySerSerGlyProAl	850
Qy	1879	GTCCCCACCCAAACCCCGCCGCGCAAGA-----GACAC	1847
Db	850	aGlyProProGlyProGlnGlyValLysGlyGluArgGlySerProGlyGlyProGlyTh	870
Qy	1846	CACAAACCGTGTAAACAGAGAGAGACTTCCCGCGCCCCCAGAGGGAGGGCGGCTGGG	1787
Db	870	rAlaGlyPheProGlyGlyArgGlyLeuProGlyProGlyAsnAsnGlyAsnProG	890
Qy	1786	CTGGCAG---CCAGTGGGCAATGGCCAAAGCGGTGCGAGCGCTGCCACATGCTAG	1730
Db	890	yProProGlyProSerGlyAlaProGlyLysAspGlyPro-----ProG	905
Qy	1729	TCCTGGGCGAGGATGGGGAAGTGGGGGGCCA-----CCTAAGAGGGAAGG	1682
Db	905	yProAlaGlyAsnSerGlySerProGlyAsnProGlyIleAlaGlyProLys--GlyAsp	924
Qy	1681	GCAGAGATAGACACAGACAGAGGGGACTGTGGAGGCCAAAGAGCCCT-----CCTCAC	1628
Db	925	AlaGlyGlnProGlyGluLysGlyProProGlyAlaGlnGlyProProGlySerProGly	944
Qy	1627	CCATC-----ATTATACAGAACCTCCACGGTTAGGA	1595
Db	945	ProLeuGlyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyProPro-Gly----	962
Qy	1594	CTGTCTTTCCCAACTCTCTACTCTCCCACTCCAGATCTCTACGTGGATCTCAC	1535
Db	963	-----MetProGlyProArgGlySerProGlyProGlnGlyIleLys-GlyGluSer--	979
Qy	1534	ATAGTCCCGCAGCACTCGCAGTTGTGCTCCAGGTAGCGGTGAGGTCTAGGTTT	1475
Db	980	-----GlyLysProGlyAlaSer--GlyHis-----A	988
Qy	1474	CTGATGAGGGGCCAGGCCCTGGGTCCCTGTGCGATTGAGAGCTGCAGGGAG	1415
Db	988	sNlglyLysGlyProPro--GlyPro-----GlnGly--	998
Qy	1414	GTGCCAGAGCACCGTGCAAGCAGCAGCTAACATCCCCACAGAGTCCCTGTGGGAGGAT	1355
Db	999	LeuProGlyGlnProGlyThrAla-----GlyGluPro	1009
Qy	1354	GGACGAGAAGCATGACGCGCGCCCGGCCCAACAGGGGAAGCAGCTGCAGCCCCCAGG	1295
Db	1010	GlyArgAspGly-----AsnProGlySerAspGlyGlnProGlyArg	1023
Qy	1294	TGTCGTCTCCCACTTGGCTGGGTGGGGGTGAGGGTGGGACACCATGAGGAGCAG	1235
Db	1024	AspGlySerPro-----GlyGlyLys	1030
Qy	1234	GGGATGGAGGGGGAACACTCTCTGCCAAGCGGGAGAGAAAGAGAGCATGTGTATG	1175
Db	1031	GlyAspArgGly-----	1034
Qy	1174	TTTGAGGATGAGGACTTGGCAGCATAGTGTCTATCAGTAGATACTTCAGAAATGATCT	1115
Db	1035	-----GluAsnGlySer	1038
Qy	1114	GTGTCTACGTGAGAGAGTCTTTGGCATGTGTGCCACTGTGTGTGCCAACATGTATAT	1055

[illegible]



QY 3973 GGCAGAGGCTCACAGCTCTCTG-----TCTCTGGCTCAACAGGCTGT-- 3931  
Db 239 yAspArgGlyLeuProGlyLeuProGlyHisLysGlyHisProGlyMetProGlyMetPro 259  
QY 3930 -GGCATACAGGCTGGCTCTCACAAGTGGAG-CAGGGTTGAAGGGGAGCGAGAGG 3873  
Db 259 oGlyMetLysGlyAlaArgGlyPheAspGlyLysAspGlyAlaLysGlyAspSerGlyAl 279  
QY 3872 AGAAGTCAAGTCACAAGCCATGAGCCCGCAGGTGAGGCTGACTGCACCTGCTGAG 3813  
Db 279 aProGlyProLysGlyGluAlaGlyGlnProGlyAlaAsnGlySer----- 294  
QY 3812 GCTGCATCTTCTTCTGAGCGGTGAAGTCTCTGGCGAGCGCCACAGCCAGGTCT-GC 3754  
Db 295 -----ProGlyGlnProGlyProGly 301  
QY 3753 AGCTCTTACAGCCAGCAAGTCTCATCTTCTGGAGAGTCACTGTGGCGAGGCCA 3694  
Db 301 yGlyProThrGlyGluArgGlyArgProGlyAsnProGlyGlyProGlyAlaHisGlyLy 321  
QY 3693 GGAGTCCAAGTGGTTCAAGTCCAGCCAGCGCTGGGCGAGTGGTAGCCAGAGCTGCC 3634  
Db 321 sAspGly-----AlaProGlyThrAlaGlyProLeuGlyProGlyProPr 337  
QY 3633 ATGACGCCCAATGTGCCCCAGCAGCGCTGTGAGGCTGTGCAGAAGTGGCGCCAGGCTG 3574  
Db 337 o---GlyThrAlaGlyPheProGlySerProGlyPheLysGlyGluAlaGlyPro---- 354  
QY 3573 CGCGGAGCTCAGCAGTGGCAGCTGAGTGGCCAGCCACCAAGTAAACAGAGGTGG 3514  
Db 355 -----ProGlyProAlaGlyAlaSerGlyAsnProGlyGluAr 367  
QY 3513 CTGTAGGCTCTAGTTCGGTTCAGCGCAGCTTGTGTCATTGAGGCTTCGCCACACCTCC 3454  
Db 367 sGlyGluProGlyProGlnGlyGlnAlaGly-----ProPr 379  
QY 3453 AAGTCAACAGTGGCTGGCAGAGTCTCTCCCCAGCGGGGAGGTTGAAGTCTGGC 3394  
Db 379 oGlyProGlnGlyProProGlyArgAlaGlySerProGlyGlyLysGly--GluMetGly 398  
QY 3393 -----TCGTTGAAGGGGGCCAGGTAGTTCAGCTGTGNAAGGAGAGGCTATGGG 3340  
Db 399 ProSerGlyLeuProGlyGlyPro-----Gly 407  
QY 3339 AAGGAGAGGCGCAGAGCGCTGCTCACCACCAAGGAGATACCTGCTCCCAAGTTCTA 3280  
Db 408 ProProGlyGlyArgGlyLeuProGlyProPro----- 418  
QY 3279 TTTTGTGTCCCTGACACTGGCCCTGTCTCCANGTAGGCTTCTTTGTTTCATGGCCTC 3220  
Db 418 ----- 418  
QY 3219 CCTGAGTTTTCATAAGGATATCATTTGTATGTCAGGTGACAGTTGAAGCAGTGGCGGT 3160  
Db 419 -----GlyThrSerGlyAsnProGly-Al 426  
QY 3159 GGTGAGCTCTCAATGAGAGGGCTGAGAAAAGGAGGAGGAGAGAC---GCACAAGAGC 3103  
Db 426 aLysGlyThrProGlyGluProGlyLys-AsnGlyAlaLysGlyAspProGlyProLysG 446  
QY 3102 CTCAGAAGAGCTGCTGATGCTGGTGGCGGAAACCTTGGGAGAGTGCCCGCTGTGGG 3043  
Db 446 lyGluArgGlyGluAsnGlyThrProGlyAlaArg-----G 458  
QY 3042 GGCAGGACAGAGAGGAGAGGCTCGAGGAGCTGGAGAG-----GTAG 2998  
Db 458 lyProProGlyGluGluGlyLysArgGlyAlaAsnGlyGluProGlyGlnAsnGlyValp 478  
QY 2997 AGGTGGCCCGAGAGTGCAGAGACCAAGGA---AGAGGAGTT----- 2957  
Db 478 roGlyThrProGlyGluArgGlySerProGlyPheArgGlyLeuProGlySerAsnGlyL 498

QY 2956 -----CGAAGAGCAGCAGCGCTCAGAGTGTGGCTGCAAAACAGACAGCAGCAAGGAGAA 2902  
Db 498 euProGlyLysGlyProAlaGlyGluArgGlySerProGlyProProGlyProSerG 518  
QY 2901 AGCTTCAAAACACTGTTCGGCTTGCAGAGGAGTCCCTACACCTTGGAGAGAACAT 2842  
Db 518 lyProAlaGlyAsp---ArgGlyGlnAspGlyGlyProGlyLeuProGlyMetArgGlyL 537  
QY 2841 TCCTGGG---CCTGGAAGCCAGATTGTCAGGGGTTAAGAGTGTAGTGGGTGGAGAGAA 2785  
Db 537 euProGlyLeuProGlySerPro-----GlyS 546  
QY 2784 GTGGAGGACAGAGGCTGTAGACCACTCTTGGAGAAATTTGGCAGGAAACCCAGGAGAGA 2725  
Db 546 erAspGlyLysProGlyProPro-----GlyAsn-----GlnGlyGluProGly 560  
QY 2724 AAAAGGTCGGGAACATAGAGGCGCAGTGGAGTGAAGCAAAAGTTTCTGCACAAGGAAC 2665  
Db 560 ----- 560  
QY 2664 AAGAGAGATGAGAGTGGCCCATTTGGAACAGAGCCCATTTGCGG----- 2618  
Db 561 -----ArgSerGlyPro-----ProGlyProAlaGlyProArGlyGlnProG 575  
QY 2617 -----CTGACACACCCAGGCCCCACCAGCT---GAGGGGCTG---GGGAGAGACGGTACT 2569  
Db 575 lyValMetGlyPheProGlyProLysGlyAsnGlyAlaProGlyLysAsnGlyGlu- 594  
QY 2568 CACTAACACTCGAAGAGGTTTCACTGATAGGACAGGTTGTCAGAAAGAAATG 2509  
Db 595 -----ArgGlyProGlyGlyProProGlyThrProGlyProAlaGlyLysAsnG 611  
QY 2508 GAGGCACTCATGATAGTAGGACAGGAGGAGGCTATGGAAGACAGCATGTATAGGGAAG 2449  
Db 611 lyAsp-ValGlyLeuProGlyProPro-----GlyProAlaGlyProAlaGlyAsp 627  
QY 2448 AGGCAGAGCCCTTTCCAAACCTCGGGGAGCGCTGGGGCCCGAGGAAAGGATGAGGG 2389  
Db 628 ArgGlyGluProGlyProSer-----GlySerProGlyLeuGlnGlyLeuProGlyGly 645  
QY 2388 AGATGTATACTGGGAAAGAGGAGGAGCAAGGAGGAGGCTGGGATAGGCAAGGCC 2329  
Db 646 ProGlyProAlaGlyGluAsnGlyLysProGlyGluProGly----- 659  
QY 2328 CAGGGGAGGTTTACACGCTGACACAGCAGACAGACAACTAATTTCTCGGTATCTCTG 2269  
Db 660 -----ProLysGlyAspIle 664  
QY 2268 GGGCTGGAGCAGCTTCCCGCCACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2209  
Db 665 GlyGlyProGlyPheProGlyProLysGlyGluAsnGlyLeuPro----- 679  
QY 2208 GTCTGGGATGTGAAGAGGACACCGGAGAGGGGGCCGGTT---CCCCATGTGTCTCG 2152  
Db 680 -----GlyGluArgGlyProGlnGlyProGlyProGlyProThrGlyAlaArg 694  
QY 2151 GCTCTCCAAACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2098  
Db 695 GlyGly-----ProGlyProAlaGlySerGluGlyAlaLysGlyProProGly 710  
QY 2097 CCCCCAAATCTTACAGATACCTCCATCCACACCTTAAACCCCTGCCACTACTGCTCT 2038  
Db 711 ProProGlyAlaProGlyGlyThrGlyLeuProGlyLeuGlnGlyMetPro-GlyGluAr 730  
QY 2037 GGGTTCCCTACAGTCCCTTCCCGCCATGACCCCTAGAGGCCCGCCAGCTAGGATGGAGA 1978  
Db 730 sGlyAlaSerGlySerProGlyProLysGlyAspLysGlyGluProGlyGlyLysGly-- 749  
QY 1977 GGCAGAGAGGAGGAGGCGCTGGGTCCAGCAGCAGGAGGCGCCCGCCAGTCTGGGGGAGAG 1918  
Db 750 -AlaAspGly-----LeuProGlyAlaArg-----GlyGluAr 760  
QY 1917 CGGAGCTGGAACGAGGCGCATGGGAGTGGGGCCCTCTCTCTCCCAACCCCGCCCGC 1858



```
Db 760 g-----GlyAsnValGlyProile-----GlyProProG 770
QY 1857 GCAAGACACCAACACACGTGTAAACAGAGAGACTTCCCGCGCCCGCCAGGGAGGC 1798
Db 770 yProAlaGlyPro-----ProGlyAspGlyGlyGluTh 781
QY 1797 GGGC-----GGCTGGCTGGCGAGCAGCGAGTGGGCGCAATGGCCAGCGCGGT 1753
Db 781 rGlyProAlaGlyAlaProGlyProAlaGlySerArgGlyGlyProGlyGluArgGlyG 801
QY 1752 GCGAGGCTGCCCATGCTAGT-----CCTGGGGCC-----AGGATGGGAAGTG 1705
Db 801 uGlnGlyLeuProGly--ProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGluPro 820
QY 1704 GGGGGGCCACCTAAGAGGGAAGGAGGAGATAGACACAGAGAGGGGAGCTGGTGGAGCC 1645
Db 821 GlyGlyLysGlyGluArgGlyProProGlyLeuArgGlyGluAlaGlyProProGlyAla 840
QY 1644 AAAGAGCCCTCTCCACCCCATCATTTTACAGAACCCCTCCACGGTTAGACTGTC----- 1590
Db 841 AlaGlyProGlnGlyGlyProGly--AlaProGlyProProGlyProGlnGlyValLysG 860
QY 1589 -----TTTCCCAACTCTCTCACTCTCTCC 1567
Db 860 yGluArgGlySerProGlyGlyProGlyAlaAlaGlyPheProGlyAlaArgGlyProPr 880
QY 1566 CCAACTCCAGATTCCTAGCTGGTGGATCTCATATAGTCCACAGCCAAAGCTGGGAGTT-- 1509
Db 880 oGlyProProGlyAsnAsnGlyAspArgGlyGluSerGlyProPro-----GlyValPr 898
QY 1508 -----GGTGTCTCCAGGTAGCGGTGAGG 1486
Db 898 oGlyProProGlyHisProGlyProAlaGlyAsnAsnGlyAlaProGlyLysAla--G 917
QY 1485 TCATAGTTTCTGATGAGAGGGCCAGCGCTGGTCCCTGCTGCTGCGATTGAGACTGGC 1426
Db 917 yGluArg-----GlyPheGlnGly--ProLeuGlyPro----- 927
QY 1425 ACTGCAGGGA-----GGTCCAGAGCACCGTGCAGGCGAC 1390
Db 928 ----GlnGlyAlaLeGlySerProGlyAlaSerGlyAlaArgGlyPro----- 942
QY 1389 GCTAATACATCCCGAGTCCCTCTGGCGAGGAGTGGAGCAGATGAGCGCGCGCCG 1330
Db 943 ----ProGlyProAlaGly--ProProGlyLysAspGlyArgGly-----GlyTyrProG 959
QY 1329 GCGCCACACGGGGAAGACAGTGGCAGCCCTCAGGTGTCTGTCTCCACCTTGGGCTGGGT 1270
Db 959 lyPro-----IleGlyProPro----- 964
QY 1269 GGGGGGTGAGGTGGGACACCATGATGGGAGGCGAGGGATGGAGGA 1223
Db 965 --GlyProArg-----GlyAsnArgGlyGluSerGly 974
RESULT 6
CA54_HUMAN STANDARD; PRT; 1685 AA.
AC P29400; Q16126; Q16006;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
```

```
RL J. Biol. Chem. 269:6008-6614 (1994).
RN [2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CVS-521.
RC TISSUE=Kidney;
RX MEDLINE=92316933; PubMed=1352287;
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481 (1992).
RN [3]
RP SEQUENCE OF 85-1685 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=90337990; PubMed=2380186;
RA Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV).";
RL J. Biol. Chem. 265:13758-13766 (1990).
RN [4]
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE=91169491; PubMed=2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9 (1991).
RN [5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610 (1990).
RN [6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";
RL Am. J. Hum. Genet. 46:1024-1033 (1990).
RN [7]
RP SEQUENCE OF 1-20 FROM N.A.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain.";
RL Kidney Int. 44:1316-1321 (1993).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499 (1997).
RN [10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18 (1991).
```



QY 1539 GTATCCAGCGTAGGAATCTGGAGTGGGGA-----GGAGTGAGGAGT----- 1581  
Db 369 -----GluysGlyGluAsgGlyPheProGlyLe 384  
QY 1582 -----TGGGAAGACAGCTCCTAACCGTGGAGGTTCTGTAATGATGGGTGA 1631  
Db 384 uProGlyProGlyAlaAvalMetGlyProProGlyProGlyPheProGlyGly 404  
QY 1632 GGAGGGCTCTTTGGCTCCACCAGTCCCTGCTGTCTCTATCTCCCTCCCTCCCTCT 1691  
Db 404 u-AsgGlyGlnLysGlyAspGluGlyProProGlyLysGlyPheProGlyL 424  
QY 1692 TAGGTGGCCCCCCTTCCCATCCCTG-----GCCCCAGGACTAGGCATGTGGGAG 1745  
Db 424 euAspGlyGlnProGlyAlaProGlyLeuProGlyProProGly----- 438  
QY 1746 GCCTGGCACCCGCTTGGCCCATTCGCCACTGGC-----TGCCAGCCAGCCGCC 1796  
Db 439 -----ProAlaGlyProHisIleProProSerAspGluIleCysGluProGlyProp 456  
QY 1797 CGCCTCCCTCCCTGGGGCGGGAGTCTCTCTGTTTACACCGTGTGTGTCTCTCTG 1856  
Db 456 roGlyProProGlySerProGlyAsp----- 464  
QY 1857 CGCGGGCGGGTGGTGGGACAGAGG----- 1885  
Db 465 -----LysGlyLeuGlnGlyGlnGlyValLysGlyAspLysGlyAspThrCysPheA 483  
QY 1886 -----GCCCACTCCCATCCCTCGTCCGTCGCTCCCTCCCTCCCTCC 1922  
Db 483 snCysIleGlyThrGlyIleSerGlyProProGlyGlnProGlyLeu----- 498  
QY 1923 CCCCAGAGCTGGGGCCCTGCTGCTGGACCCAGGGGCTCCCTCCCTCCCTCCCTCC 1982  
Db 499 -----ProGly-----LeuProGlyProProGlySerLeuGlyPheP 511  
QY 1983 CATCTAGCTGGGCTCTCTAGG-----GGGTATGGGG-----AAGGGAGCTAGGGA 2033  
Db 511 roGlyGlnLysGlyGluLysGlyGlnAlaGlyAlaThrGlyProLysGlyLeuProGlyI 531  
QY 2034 ACCAGGAGTAGTGGCAGG-----GGGTTAGGTTGGATGGAGGTTATGCTGTAAAGAT 2090  
Db 531 leProGlyAlaProGlyAlaProGlyPhePro-----Glys 543  
QY 2091 TTGGGGGTGGTCCA--GAGGTGTTTCAGAGACCCAGGAGAGAGAGAGG-----G 2141  
Db 543 erLysGlyGluProGlyAspIleLeuThrPheProGlyMetLysGlyAspLysGlyGluL 563  
QY 2142 TTGGAGGAGCCGAGCA-----CCATGGGGAACCGGCCCTCTCTCC 2183  
Db 563 euGlySerProGlyAlaProGlyLeuProGlyLeuPro-GlyThrProGlyGlnAspGly 582  
QY 2184 CGTGTCTCTTCCATCCAGACCTACTCTGGAGCCAGGAGAGAGAGAGAGAGAG 2243  
Db 583 LeuProGlyLeuProGlyProLysGlyGluProGlyGlyIleThrPheLysGlyGluArg 602  
QY 2244 -----GTGGCGGGGAGCTGGCTCA-----GCC 2267  
Db 603 GlyProProGlyAsnProGlyLeuProGlyLeuProGlyAsnIleGlyProMetGlyPro 622  
QY 2268 CCAGGATACCGAGGAAATAGTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 2327  
Db 623 ProGlyPhe-----GlyProPro 628  
QY 2328 GGGCCCTTGCCTATCCAGGCTCTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2387  
Db 629 GlyProValGlyGluLysGlyIleGlnGlyValAlaGlyAsnProGlyGlnProGlyIle 648  
QY 2388 TCCCTCATCCCTTCTCCCTGGGCC-----CAGCGGCTC 2420  
Db 649 -----ProGlyProLysGlyAspProGlyGlnThrIleThrGlnProGly 663  
QY 2421 CCCCAGGGGTG-----GAAAGG 2438

Db 664 LysProGlyLeuProGlyAsnProGlyArgAspGlyAspValGlyLeuProGlyAspPro 683  
QY 2439 GCTCTGGCCCTTCTCCATATACCATGCTCTTCCATAGCCCTTCTCTCTCTCTCTCT 2498  
Db 684 GlyLeuProGlyGln-----ProGlyLeuProGlyIleProGlySer 697  
QY 2499 AGACTGCTCCATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2558  
Db 698 LysGlyGluProGlyIleProGlyIleGlyLeu-----ProGlyProPro 712  
QY 2559 AGTGTAGTAGTACCGCTCTCTCCAGCCCTCAGCTGCTGGCTGGCTGGCTGGCTGCT 2618  
Db 713 GlyProLysGlyPheProGlyIleProGlyPro-----ProGlyAlaProGly---ThrPro 730  
QY 2619 GGCATAATGGGCG-----TCTGTTTCCAAATGGGCGCACTCTCATCT 2657  
Db 731 GlyArgIleGlyLeuGluGlyProProGlyProProGlyPheProGlyProLysGlyGlu 750  
QY 2658 CTCTCTTGTCTTGTGTCAGAAAACCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2717  
Db 751 ProGlyPheAlaLeuProGlyProPro-----GlyProProGlyLeuProGlyPheLys 768  
QY 2718 CCGTTTCTCTCTCT-----GGCTTCTCTCCCAATTTCTCTCA-----AGG 2759  
Db 769 GlyAlaLeuGlyProLysGlyAspArgGlyPheProGlyProProGlyProGlyArg 788  
QY 2760 AGTGTCTTACACCTCTCTCT-----CGACTT 2786  
Db 789 ThrGlyLeuAspGlyLeuProGlyProLysGlyAspValGlyProAsnGlyGlnProGly 808  
QY 2787 CTTCTCCACCCACTCTTCTTAAACCCCTTCAATCTGGCTTCCAGGCCCCAGCAATGT 2846  
Db 809 ProMetGlyPro-----ProGlyLeuProGlyIle 818  
QY 2847 TCTCTCAGGTGCTCAGGACCTCTTCCAGCCGACAGCTGTTTGAAGGCTCATTC 2906  
Db 819 GlyValGlnGlyProProGlyProPro-----GlyIleProGlyProIle 833  
QY 2907 TCCTTGTCTGTCTTGTGTCAGCCACTGCTGAGCGTCTGCTCTCTCGAACTCTCTCT 2966  
Db 834 Gly-----GlnProGlyLeuHisGlyIle 841  
QY 2967 CTTTGTCTCTGCACTCTCTGGCCACCTTCTTACTCT---CCAGCTCTCCAGGCTCC 3023  
Db 842 ProGlyGluLysGlyAspProGlyProGlyLeuAspValProGlyProProGlyGlu 861  
QY 3024 TCTTCT 3083  
Db 862 ArgGlySerProGlyIlePro-----GlyAlaProGlyProIleGlyProProGlySer 879  
QY 3084 TCAGCACGCT 3113  
Db 880 ProGlyLeuProGlyLysAlaGlyAlaSerGlyPheProGlyThrLysGlyGluMetGly 899  
QY 3114 -----TCT 3158  
Db 900 MetMetGlyProProGlyProProGlyProLeuGlyIlePro----- 913  
QY 3159 CACCCCACTGCTTCACTGTCTCACCTGTATATATATATATATATATATATATATAT 3218  
Db 913 ----- 913  
QY 3219 GGAGGCCATGAACAAAGAGCCCTAGCATGGAGACAGGCGCAGTGTCTAGGGGACACAAAA 3278  
Db 914 -----GlyArgSerGly 917  
QY 3279 ATAGAACTTTTGGAGCAGGTATCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3338  
Db 918 ValProGlyLeuLysGlyAspGlyLeuGlnGlyGlnProGlyLeuProGlyPheProThr 937  
QY 3339 TCCCATCACCT 3398





QY 3128 TTTTCTACGCTCTCCA-----TTGGAGAGCTCACACCGCCACTGCTTCAACTGTCA 3181  
D 381 rPheSerProPro--ProProThrTyrGluGlnSerProProProProProAlaTyrSer 400  
QY 3182 CCTGCATCAAAATGATATCTTATTGGAATACTCAGGAGGCGCATGAAACAAGAGCCT 3241  
D 401 ProProLeuProAlaProProProThrTyr-----SerProProProProThrTyrSerPro 418  
QY 3242 AGCATGGAGACA-----GGCCAGTGTTCAGGGGACACAAAAATAGAACTTT 3289  
D 419 ProProProThrTyrAlaGlnProProProLeu----- 429  
QY 3290 GGGAGCAGTATCTCTTGTGTGTAGCCAGCGCTCTGCCCTCTCTCTCT-----TC 3340  
D 430 -----ProProThr--TyrSerProProProProProAlaTyrSe 441  
QY 3341 CCCATCACCTCTCTCTTTTTCACAGCTGAATCACTGGGCGCCCTTTCAAGAGCCAGAC 3400  
D 441 rProProProProPro-----ThrTyrSerProProProProProTyrSe 456  
QY 3401 TTCAACCTCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCACTGTGTGACTTGGAGGTG 3460  
D 456 rProProProProAlaTyrAlaGlnProProProProPro----- 470  
QY 3461 TGCGGAAGCTCAATGACAAACTCGCGCTGACCCAGAACTACGAGGCGCTACAGCACCTT 3520  
D 471 -----ThrTyrSerProProProProProAlaTyrSerPro--ProProPro 485  
QY 3521 CTGTGTACTTGGTGGCTCAACCGCTCAGCTGAGTGCCTGAGCTGGCGGAGGCTG 3580  
D 485 eProPheTyrSerProProProProGln-----ValGlnProL 498  
QY 3581 GCCACTTTCGACAGCTCCAGGGCGCTGCGGAGCATTTGGGCGCTCATGGCAGCT 3640  
D 498 euProProThrPheSerProProProProProProArgGlnHis----- 511  
QY 3641 CTGGGTACCACTGGCC-----CAGCCGCTGCTGGGACT----- 3676  
D 512 -----LeuProPro--ProProHisArgGlnProArgProProProThrTyrGlyGln 529  
QY 3677 -----GAACCACTTGGACTCTGGGCT-----GCCACAGTACTTCTCT 3718  
D 530 ProProSerProProProPheSerProProProProProArgGlnHisSer--ProProPr 549  
QY 3719 CAGAAGATGAGAGCTTCTGGCTGTCTGAGGAGCTGAGAGCTGGCTGGCTGGCC 3778  
D 549 oProHisTyrGlnPro-----ArgThrProProProProThrTyrGlyGlnProPr 565  
QY 3779 AAGGACTTCAACGGCTCAAGAGAAGATGAGCTCCAGCAGCTGAGTCAACCTGCAC 3838  
D 565 oSerProProThrPheSerAlaPro-----ProProArgGlnHisSerPro----- 581  
QY 3839 CTGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3898  
D 582 -----ProProProHisArgGln--ProArgProProPro 592  
QY 3899 CTCCCACTTGTGAGAGCCGCTGTATGCCAACACCTGTGAGGCGAGGAGAGCAGAGC 3958  
D 592 hrProThrTyrGlyGlnPro--ProSerProProProProThrTyrSerPro----- 607  
QY 3959 TGTGAGCCTCTGGCCCTTCTGGACCGGCTGGCGGTGTGATCGGATCAGCCCTGTCTCC 4018  
D 608 -----Pro 608  
QY 4019 TCCCACTTCCCAAGGTCTA 4039  
D 609 SerProProProTyrGlyLeu 615  
RESULT 8  
BAT2\_HUMAN  
ID BAT2\_HUMAN  
AC P48634,  
STANDARD; PRT; 2142 AA.

DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Large proline-rich protein BAT2 (HLA-B-associated transcript 2).  
GN BAT2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBT\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=T-cell;  
RX MEDLINE=90192810; PubMed=2156268;  
RA Banerji J., Sands J., Strominger J.L., Spies T.;  
RT "A gene pair from the human major histocompatibility complex encodes  
RT large proline-rich proteins with multiple repeated motifs and a  
RT single ubiquitin-like domain."  
RL Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).  
RN [2]  
RP SEQUENCE OF 1-1860 FROM N.A.  
RX MEDLINE=93272029; PubMed=8499947;  
RA Iris F.J.M., Bougueleret L., Prieur S., Caterina D., Primas G.,  
RA Perrot V., Jurka J., Rodriguez-Tome P., Claverie J.-M., Dausset J.,  
RA Cohen D.;  
RT "Dense Alu clustering and a potential new member of the NF kappa B  
RT family within a 90 kilobase HLA class III segment."  
RL Nat. Genet. 3:137-145(1993).  
CC -1- FUNCTION: UNKNOWN.  
CC -2- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; M33509; AAA35585.1; -;  
DR EMBL; M33518; AAA35586.1; -;  
DR EMBL; M33512; AAA35586.1; JOINED.  
DR EMBL; Z15025; CAA78744.1; -;  
DR PIR; B35098; B35098.  
DR PIR; S36152; S36152.  
DR Genew; HGNC:13918; BAT2.  
DR TM; 142580; -;  
KW Repeat.  
FT DOMAIN 519 524 POLY-PRO.  
FT DOMAIN 636 657 GLN-RICH.  
FT DOMAIN 684 688 POLY-PRO.  
FT DOMAIN 699 704 POLY-PRO.  
FT DOMAIN 814 821 POLY-PRO.  
FT DOMAIN 1340 1345 POLY-GLY.  
FT DOMAIN 1398 1403 POLY-GLY.  
FT DOMAIN 1436 1442 POLY-PRO.  
FT DOMAIN 1982 1991 POLY-PRO.  
FT DOMAIN 41 1795 4 X 57 AA TYPE A REPEATS.  
FT REPEAT 41 95 1-1.  
FT REPEAT 98 154 1-2.  
FT REPEAT 281 337 1-3.  
FT REPEAT 1740 1795 1-4.  
FT DOMAIN 337 549 2 X TYPE B REPEATS.  
FT REPEAT 337 418 2-1.  
FT REPEAT 476 549 2-2.  
FT DOMAIN 1899 2089 3 X 50 AA TYPE C REPEATS.  
FT REPEAT 1899 1948 3-1.  
FT REPEAT 1965 2014 3-2.  
FT REPEAT 2040 2089 3-3.  
FT CONFLICT 57 57 R -> A (IN REF. 2).  
FT CONFLICT 109 109 Q -> S (IN REF. 2).  
FT CONFLICT 414 414 P -> PPHRGFAGNAGPP (IN REF. 2).  
FT CONFLICT 532 532 T -> K (IN REF. 2).  
FT CONFLICT 682 682 Q -> K (IN REF. 2).

FT CONFLICT 730 730 E -> D (IN REF. 2).  
 FT CONFLICT 750 750 L -> R (IN REF. 2).  
 FT CONFLICT 834 834 A -> T (IN REF. 2).  
 FT CONFLICT 1035 1035 G -> A (IN REF. 2).  
 FT CONFLICT 1068 1068 M -> L (IN REF. 2).  
 FT CONFLICT 1285 1285 P -> R (IN REF. 2).  
 FT CONFLICT 1400 1400 G -> A (IN REF. 2).  
 FT CONFLICT 1611 1611 T -> S (IN REF. 2).  
 FT CONFLICT 1729 1729 G -> A (IN REF. 2).  
 SQ SEQUENCE 2142 AA; 227840 MW; 32DDF16B9B52420A CRC64;  
  
 Alignment Scores:  
 Pred. No.: 2,29e-07 Length: 2142  
 Score: 313.50 Matches: 477  
 Percent Similarity: 29.88% Conservative: 178  
 Best Local Similarity: 21.76% Mismatches: 670  
 Query Match: 3.32% Indels: 869  
 DB: 1 Gaps: 119  
  
 US-09-931-704-3 (1-5087) x BAT2\_HUMAN (1-2142)  
 QY 472 TGCCACATCCCTCTTACAGCACCAACCTGGCTCTGGACTCTGGTCTCTGGAT 531  
 DB ||||| |||  
 DB 39 CysProSerPro-----TriPro 44  
 QY 532 GTCCAAACTCTGCAG---TGCCATCAGCAACAAGCCGACTCGTCAAATGCACCTCTCT 588  
 DB ||||| |||  
 DB 45 AlaGluSerArgGluSerCysHisCysProAla-----TyrArgPro 58  
 QY 589 CCCTTCCTGTGCCACCTCTGCAGCTGATGGAAGGCTCATTTGAAGTCCAACTTTCC 648  
 DB ||||| |||  
 DB 59 ProAlaAsnLeuProSerLeuLysAlaGluAsnLysGly-----Asn 72  
 QY 649 CCACCTAACACC-----AAGACGGGGTGAACCTCCACACTCCACCTCCCTCC 696  
 DB ||||| |||  
 DB 73 AspProAsnValSerLeuValProLysAspGlyThrGlyTrp----- 86  
 QY 697 TGAGAGTGAGCACTAAATCTCTTCAATCTTAACCCACCCCTACACTTCCACACTCAGGA 756  
 DB ||||| |||  
 DB 87 -----AlaSerLysGlnGluGlnSerAsp----- 94  
 QY 757 ATCATCATCTAGATATATACCAAACTAAGCCCCNATAAGCGCCGACCTAGTGGTCT 816  
 DB ||||| |||  
 DB 95 -----ProLysSerSerAspAlaSerThrAlaGlnProProGluSer 108  
 QY 817 AACCTATACCTTGTCTCTATGGGTGAGTCTGTTCTTGGCGCGCCCTCTCTCTGCTT 876  
 DB ||||| |||  
 DB 109 GlnProLeuPro-----AlaSerGln--ThrProAla 119  
 QY 877 CCTCCTTAGAGCTGACTGTGCTCAGCTTGCAGCTCTGACATGTGCTCTCTCCACCCCT 936  
 DB ||||| |||  
 DB 119 eAsn-----GlnProLysArgProProAlaAla-----ProG 130  
 QY 937 CTGACTCCCTCAAGCTGAGTGGGACTGGAGACTGGCAGGAGCTAGGTA----- 989  
 DB ||||| |||  
 DB 130 luAsnThrProLeuValProSerGlyValLysSerTrpAlaGlnAlaSerValThrHis 150  
 QY 989 ----- 989  
 DB 150 lyAlaHisGlyAspGlyArgAlaSerSerLeuLeuSerArgPheSerArgGluGluP 170  
 QY 990 -----CAACTGGAACACAGGCGAGGTGCACC 1014  
 DB ||||| |||  
 DB 170 heProThrLeuGlnAlaAlaGlyAspGlnAspLysAlaLysGluArgGluSerAlaG 190  
 QY 1015 TCGAGTCCCTAGCCTGGCCGCTCCCTCCATGATACACATATACATATGTTGGCACACAC 1074  
 DB ||||| |||  
 DB 190 luGlnSerSerGlyProGlyProSerLeuArgProGlnAsnSerThrThrTrpArgAspG 210  
 QY 1075 ACAGTGGCACACATGCCAAAGACTCTCTCAGCT---GACACACAGATCATCTCAAGTA 1131  
 DB ||||| |||  
 DB 210 lyGlyGly-ArgGlyProAspGluLeuGluGlyProAspSerLysLeuHis---HisGly 228

QY 1132 TCTACTGATAGACACTCATCGTGCACAGTCTCTCATCTCAAAATACATACATGCTCTCT 1191  
 DB ||||| |||  
 DB 229 HisaspProArgGlyGlyLeuGlnProSerGlyProGln----- 242  
 QY 1192 TTCTCTCCCTGTTGCCAGGAGTGTTCCTCTCTCTCCCTCTCCCTCTCCCTCTCCCTCT 1251  
 DB ||||| |||  
 DB 243 PheProProTrpArgGlyMetMetProPheMetTyrProProTyrLeuProPhe--- 261  
 QY 1252 GTCCACCTCATCCGCCCCACCCAGTGGGGGACACACCTGAGGGGCTGCCAGC 1311  
 DB ||||| |||  
 DB 262 ProPro-ProTyrGlyProGlnGlyProTyrArgTyrProThrProAsp-GlyProSerA 281  
 QY 1312 TGCTTCCCCGTGTG-----GCCCG----- 1331  
 DB ||||| |||  
 DB 281 rgPheProArgValAlaGlyProArgGlySerGlyProProMetArgLeuValGluProV 301  
 QY 1332 --GCCCG----- 1337  
 DB ||||| |||  
 DB 301 alGlyArgProSerIleLeuLysGluAspAsnLeuLysGluPheAspGlnLeuaspGlnG 321  
 QY 1338 -----GCTCATGCT----- 1346  
 DB ||||| |||  
 DB 321 luAsnAspAspGlyTrpAlaGlyAlaHisGluGluValAspTyrThrGluLysLeuLysP 341  
 QY 1347 -----TCTCTCATCTCTCCACAGGAGCTCTGGGGATGTTAGC 1389  
 DB ||||| |||  
 DB 341 heSerAspGluGluaspGlyArgAspSerAspGluGluGlyAlaGluGlyHis-----A 359  
 QY 1390 GTGCTGTGCAGGTGCTCTGGCACCTCCCTGCAGTGCAGCTCTCAATCGCACAGGGA 1449  
 DB ||||| |||  
 DB 359 rgAspSerGlnSerAlaSerGlyGluGluArgProProGlu-----AlaaspGlyL 376  
 QY 1450 CCCAGGCTCGCCCT-----CCAT 1470  
 DB ||||| |||  
 DB 376 ysLysGlyAsnSerProAsnSerGluProProThrProLysThrAlaTrpAlaGluThrs 396  
 QY 1471 CCAGAAACCTATGACCTACCTACCTGAGCACCACTCCCGAGCTT----- 1521  
 DB ||||| |||  
 DB 396 erArgProProGluThrGluProGly-ProProAlaProLysProProLeuProProGly 415  
 QY 1522 -----GGCTGGGACCTATGTAGTATCCAGCG-----TAGGAATCTGGG 1560  
 DB ||||| |||  
 DB 416 AspTyrProAspArgGlyGlyProProCysLysProProAlaProGluAspGluGlu 435  
 QY 1561 AGTTGGGAGGAGTGAAG-----AGTTGGGAAAGACAGTCTTAACCTGAGGGTCT 1614  
 DB ||||| |||  
 DB 436 AlaTrpArgGlnArgLysGlnSerSerSerGluIleSerLeuAlaValGluArgAla 455  
 QY 1615 GGTAAATGATGGGTGAGGAG----- 1635  
 DB ||||| |||  
 DB 456 ArgArgArgGluGluGluArgMetGlnGluGluArgMetGlnGluGluArgAlaCysAla 475  
 QY 1636 -----GGGCTCTTTGGCTCCACAG----- 1656  
 DB ||||| |||  
 DB 476 GluLysLeuLysArgLeuAspGluLysPheGlyAlaProAspLysArgLeuLysAlaGlu 495  
 QY 1657 -----TCCCCCTGTCTGTCTATCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1710  
 DB ||||| |||  
 DB 496 ProAlaAlaProProAlaAlaProSerThrProAlaPro-----ProProAlaVal 512  
 QY 1711 CCC-----ATCCCTGGCCCGAGGAGTGGGCGAGGCTGGGCGAGGCTGGCCTGGC 1764  
 DB ||||| |||  
 DB 513 ProLysGluLeuProAlaProProAla-----ProProAlaSerAla 527  
 QY 1765 CCATTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1824  
 DB ||||| |||  
 DB 528 ProThrProGluThrGluProGluGluProAlaGlnAlaPro-----ProAlaGlnSer 545  
 QY 1825 CCTCTGTTTACACCTGTTGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1878  
 DB ||||| |||  
 DB 546 -----ThrProThrProGlyValAlaAlaAlaAlaProThrLeuValSerGlyGly 562  
 QY 1879 CAGAGGGGGCCACCTCCCATGCTGCT-----GTTCCA 1911

Db 563 SerThrSerSerThrSerSerGlySerPheGluAlaSerProValGluProGlnLeuPro 582  
::: |||  
QY 1912 GCTCGCTCTGCCCCAGACCTGGGGCCCTCTGCTCTGACCCAGGGGCTCCCTTCG 1971  
::: |||  
Db 583 SerLysGluGlyProGluPro-----ProGluGluValProPro-----Pro 596  
::: |||  
QY 1972 TCTGCTCTCTCCATCTAGCTGGCCCTCTAGGGGGTCTAGGGGGAGGGGACTGTAGG 2031  
::: |||  
Db 597 ThrThrProProValProLysValGluProLysGly---AspGlyIleGlyProThrArg 615  
::: |||  
QY 2032 GAACCCAGGCAGTAGTGGCCAGGGGTTTAGGGTCTGGATGGAGGTATTGCTGAAGGATT 2091  
::: |||  
Db 616 GlnProProSer-----GlnGlyLeu-----GlyTyrProLysTyrGln 628  
::: |||  
QY 2092 TGGGGTGGTCCAGAGTGTTCAGAGCCAG-----GAGAGAGG 2133  
::: |||  
Db 629 LysSerLeuProProArgPheGlnArgGlnGlnGlnGlnGlnLeuLeuLysGlnGlnGln 648  
::: |||  
QY 2134 AAGGAGGGTTGGAGGAGCCGAGGACCACATGGGGAACCGGCCCTCTT----- 2181  
::: |||  
Db 649 GlnHisGlnTrpGlnGlnHis---GlnGlnGlySerAlaProProThrProValProPro 667  
::: |||  
QY 2182 -----CCGGTTCCTCTCCACATCCAGACCTACTCTGGAGCCAGGAAA 2229  
::: |||  
Db 668 SerProProGlnProValThrLeuGlyAlaValProAlaProGlnAlaProProPro 687  
::: |||  
QY 2230 GAAAG-----GGAAGAAGTGGGGGGAGCTGGCTCCAGCCCGAGAT--- 2274  
::: |||  
Db 688 ProLysAlaLeuTyrProGlyAlaLeuGlyArgProProProMetProMetAsnPhe 707  
::: |||  
QY 2275 ACACCGAGGAAATTAGTTGCTCTGTGCTGTGTGAGCGGTGAACCTCCCTC-----CTG 2328  
::: |||  
Db 708 AspProArgTrpMetMet-----IleProProTyrVal 718  
::: |||  
QY 2329 GGCCTTGCTATCCAGGCTCTCCCTTGCTCTCCCTCTCTTCCAGGTATACATCT 2388  
::: |||  
Db 719 AspProArgLeuLeuGlnGlyArgPro---ProLeuGluPheTyrProProGlyValHis 737  
::: |||  
QY 2389 CCTCATCTCCCT-----TCCCTGGGCCCGCCGCTCCCTCC----- 2424  
::: |||  
Db 738 ProSerGlyLeuValProArgGluArgSerAspSerLeuGlyLeuSerGluProPhe 757  
::: |||  
QY 2425 -----GAG 2427  
::: |||  
Db 758 AspArgHisAlaProAlaMetLeuArgGluArgGlyThrProProValAspProLysLeu 777  
::: |||  
QY 2428 GGTGGAAAGGCTCTGCCCTCTTCCCTATACCATGCTGTCTTCCATAGCTTCTCTCTG 2487  
::: |||  
Db 778 AlaTrp-ValGlyAspValPheThrAlaThrPro---AlaGluProArgProLeuThrSe 796  
::: |||  
QY 2488 TCTACTCATGAGACTGCC-----TCCATTCTTCTCTCTGCAACCTGCTC 2532  
::: |||  
Db 796 rProLeuArgGlnAlaAlaAspGluAspLysGlyMetArgSerGluThrProProVa 816  
::: |||  
QY 2533 TCTATCAGCTGACCTCTCTTCCGAGCTTAGTAGTACCGCTCTCTCCAGGCCCTC 2592  
::: |||  
Db 816 lPro-ProProProProTyrLeuAlaSer-----TyrProGlyPheProGluAsnG 833  
::: |||  
QY 2593 CAGCTGGTGG-----CTGGGTGTGTCAGCGCA 2622  
::: |||  
Db 833 lYalaProGlyProProLysSerArgPheProLeuGluGluProGlyProArgProLeuP 853  
::: |||  
QY 2623 AATGGGCTCTGTTCCAAATGGGCCACTCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 2682  
::: |||  
Db 853 roTrpProProGlySerAspGluValAlaLysIleGlnThrProProProLysLysGluP 873  
::: |||  
QY 2683 CTTTGTCTTCACTCCACTGCC-----CTCTCTAGTT 2712  
::: |||  
Db 873 roProLysGluGluThrAlaGlnLeuThrGlyProGluAlaGlyArgLysLeuProAlas 893  
::: |||  
QY 2713 CCCGA----- 2717  
::: |||

Db 893 erArgSerGlyAlaGlyProProProProArgArgGluSerArgThrGluThrArgTrpG 913  
::: |||  
QY 2718 -----CCCTTTTCTCTCTGCTGCTTTTCCCTGCC 2745  
::: |||  
Db 913 lyProArgProGlySerSerArgGlyIleProProGluGluProGlyAlaProProA 933  
::: |||  
QY 2746 AAATTTCT-----CCAAGGAGTGGTCTACACCTCTCTCTCTCTCTCTCTCTCTCT 2784  
::: |||  
Db 933 rArgAlaGlyProIleLysLysLysProProProProLysValGluGluLeuProProL 953  
::: |||  
QY 2785 TTCCTCTCCAC-----CCACTCACTTCTTAACCCCTCCAACTCTGGCTTCCAGGCCCC- 2837  
::: |||  
Db 953 ysProLeuGluGlnGlyAspGluThrProLysProProLys-----ProAspProL 970  
::: |||  
QY 2838 -----AGCAATGGTCTCTCCAAGGT---CGTCAGGCACCTCTCT- 2873  
::: |||  
Db 970 euLysIleThrLysGlyLysLeuGlyProLysGluThrProProAsnGlyAsnLeuS 990  
::: |||  
QY 2874 -----TGCC 2877  
::: |||  
Db 990 erProAlaProArgLeuArgArgAspTyrSerTyrGluArgValGlyProThrSerCysA 1010  
::: |||  
QY 2878 AAGCCCGA----- 2885  
::: |||  
Db 1010 rGlyArgGlyArgGlyGluTyrPheAlaArgGlyArgGlyPheArgGlyThrTyrGlyG 1030  
::: |||  
QY 2886 -----CAGTGTTTTGA- 2897  
::: |||  
Db 1030 lyArgGlyArgGlyGlnAlaAsnSerAlaValThrGluSerPheGluGluMetMetG 1050  
::: |||  
QY 2898 -----GGCTCATCTCTCTGCTGTTTGGAGCCACACTGCTGAGCCCTGC 2946  
::: |||  
Db 1050 lyValGluValGlyGlnGlyAspGlnThrThrLeuLeuLeuProGlu-----AlaAlaM 1068  
::: |||  
QY 2947 TGCCTTCT----- 2954  
::: |||  
Db 1068 etProAlaArgHisGlyAlaArgValGlnSerMetArgLysSerProSerGlyAlaGlys 1088  
::: |||  
QY 2955 -----C 2955  
::: |||  
Db 1088 erGlyAlaGlnLysGlnAlaAlaArgProMetArgValIleTrpLeuLeuGlnThrArgA 1108  
::: |||  
QY 2956 GAACCTCTCTCTCTGCTCTCTGCACTCTCTCT-----GGGCCACCTCTTACCT 3003  
::: |||  
Db 1108 rGluProHisProArgArgGluHisSerProArgSerSerArgSerProThrThrArgS 1128  
::: |||  
QY 3004 CTCAGCTCTCCAGGCT 3059  
::: |||  
Db 1128 erProThrLeuHisArgAlaProAlaArgPheThrCysProGlyValGlyGluSerSerL 1148  
::: |||  
QY 3060 -----CCCAAGTTTGC- 3071  
::: |||  
Db 1148 euProGluGlyAlaIleSerProGlyProArgArgArgGluAlaProProGlnValCysP 1168  
::: |||  
QY 3072 -----CCACCCAGCCAATCAGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3120  
::: |||  
Db 1168 roGlyTyrSerProProAlaLysSerLeuAlaProLys-----LysProP 1183  
::: |||  
QY 3121 CT-----CCTCTTTTCTAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3168  
::: |||  
Db 1183 roThrGlyProLeuProProSerLysGluProLeu----- 1194  
::: |||  
QY 3169 GTTCAACTGTCTACCTGCATACAAATGATATCTCTATT- 3206  
::: |||  
Db 1195 -----LysGluLysLeuIleProGlyProLeuSerProValA 1207  
::: |||  
QY 3207 -----GGAAAACTCAGGGCCATGAAACAAAGAGCCTAGCATGGAGACAGGCCACT 3261  
::: |||  
Db 1207 laArgGlyGlySerAsnGlyGly-SerAsnValGlyMetGluAspGlyGluArgProArg 1226  
::: |||  
QY 3262 GTACAGGGACAC-----AAAAAATAGAAA 3285  
::: |||  
Db 1227 ArgArgArgHisGlyArgAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1246  
::: |||



QY 3286 CTTTGGAGCAGGTATCTCTTGTGTGTGAGCCAGCGCTCT-----GCCCTC 3333  
Db 1247 GlnGluArgGluAsnAla-----AlaArgGlySerGluGlyLysProSerLeu 1262  
QY 3334 CTCCTTCCCATCACCTCTCTTTTCCAGCTCAACTACCTGGGCC-----CCCTTT 3387  
Db 1263 ThrLeuPro-----AlaSerAlaProGlyProGluGluAlaLeu 1275  
QY 3388 CAAGCAGCCAGACTTCAACCTCCCGCT-----GGG 3420  
Db 1276 ThrThrValThrValAlaProAlaProProArgAlaAlaLysSerProAspLeuSer 1295  
QY 3421 GGCAGAGACTCTGCCAGGCCACTGTGTGACTTGGAGTGTGGCAGCCCTCAATGCAA 3480  
Db 1296 AsnGlnAsnSerAspGln-----AlaAsnGluGluTrpGlu 1307  
QY 3481 ACTCGGCTGAC-----CCAGAACTACGAGCCCTACAG-----CCACT 3519  
Db 1308 ThrAlaSerGluSerAspPheThrSerGluArgArgGlyAspLysGluAlaProPro 1327  
QY 3520 TCTGTGTACTT-----GCTGGCCTCAACGCTCAGCTGCCACTGCTGAGCTCG 3570  
Db 1328 ProValLeuLeuThrProLysAlaValGlyThrProGlyGlyGlyGlyGlyAlaVal 1347  
QY 3571 CCG-----GGGCTA 3648  
Db 1348 ProGlyIleSerAlaMetSerArgGlyAspLeuSerGlnArgAlaLysAspLeuSerLys 1367  
QY 3574 -----CAGCTGGC---CCACTTTCGACACCTCCAGGGCTGCT 3612  
Db 1368 ArgSerPheSerGlnArgProGlyMetGluArgGlnAsnArgArgProGlyProGly 1387  
QY 3613 GGGCAGCATTCGGCGCTCAGGAGCTCT-----GGGCTA 3648  
Db 1388 GlyLys---AlaGlySerSerGlySerSerSerGlyGlyGlyGlyGlyProGlyGly 1406  
QY 3649 CCCACTGCCAGCGCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCCA--- 3705  
Db 1407 ArgThrGlyProGlyArgGly-----AspLysArgSerTrpProSerProLys 1422  
QY 3706 ---CAGTGACTTCCTCCAGAGAT-----GAGCAGCTTCTGGCTGCTGAGGAGCT 3753  
Db 1423 AsnArgSerArgProGluGluArgProGlyLeuProLeuProProProPro 1442  
QY 3754 GCAGACTGGCTGTG---GCGCTCGGCCAA-----GGACTCAA 3789  
Db 1443 SerSerSerAlaValPheArgLeuAspGlnValIleHisSerAsnProAlaGlyIleGln 1462  
QY 3790 CCGGCTCAAGAGAGATGCGACCTCCAGCAGCTCGACTCACCTGCACCTGGGGGCTCA 3849  
Db 1463 GlnAlaLeuAlaGln---LeuSerSerArgGlnGlySerValThrAlaProGlyGlyHis 1481  
QY 3850 TG-----CCTTCT 3857  
Db 1482 ProArgHisLysProGlyProGlnAlaProGlnGlyProSerProArgProThr 1501  
QY 3858 GACTTCTGACCTTCTCTCTCTGCTCCCTTCAAACTCTCCACTTCTGAGAGCC 3917  
Db 1502 ArgTyrgluProGlnArgValAsnSerGlyLeuSerSerAspProHisPheGluGluPro 1521  
QY 3918 AGCCCTGTATGCCAACCTGTGTGAGCCAGGAGACAGAGCTGTGAGCTCTGGCCCTTT 3977  
Db 1522 GlyProMetValArgGlyValGlyGlyThrProArgAspSerAlaGlyValSerProPhe 1541  
QY 3978 CTGACCGGCTGGGGT-----GTGATCGATCAGCC 4010  
Db 1542 Pro---ProLysArgArgGluArgProProArgLysProGluLeuLeuGlnGluSer 1560  
QY 4011 CTGTCTCTCCCACTCCCAAGGTCTACCGAGCTGGGAGAGGTACAGTAGGCCCTG 4070  
Db 1561 LeuProProProHisSerSerGlyPheLeuGlySerLysProGlyProGln 1580

QY 4071 TCCTGTCTCTTCTTACAGGAAGTCACTGCTCGAGGGAGTGTGAAGTGTTCAGGTG--- 4127  
Db 1581 AlaGluSerArgAspThrGlyThrGluAlaLeuThrProHisIleTrpAsnArgLeuHis 1600  
QY 4128 -----GTGCAGAGCGCTCATGCGCTCTGCTCTTCTGCTACCA----- 4166  
Db 1601 ThrAlaThrSerArgLysSerTyArgProThrSerMetGluProTrpMetGluProLeu 1620  
QY 4166 ----- 4166  
Db 1621 SerProPheGluAspValAlaGlyThrGluMetSerGlnSerAspSerGlyValAspLeu 1640  
QY 4167 -----CTTGCCAGTGC----- 4178  
Db 1641 SerGlyAspSerGlnValSerSerGlyProCysSerGlnArgSerSerProAspGlyGly 1660  
QY 4179 -----CCACCCAGCCCTCAGTGGCAGCATCTGG-----AGG 4210  
Db 1661 LeuLysGlyAlaAlaGluGlyProProLysArgProGlyGlySerSerProLeuAsnAla 1680  
QY 4211 CGAGGGTGTAGGGGCCA-----CCACCAACATGCTCTTC 4246  
Db 1681 ValProCysGluGlyProProGlySerGluProProArgArgProProAlaProHis 1700  
QY 4247 TGGGTGAAGCCCTTGGCTGCCCTCTCTGATGGGTGTCTCCCTTATCCCA 4306  
Db 1701 AspGlyAspArgLysGluLeuProArgGluGlnProProGlyProGlyGlyThr 1720  
QY 4307 AATCACTCTATACATCCAAATTCAGAAACAAACATGTGGCAATTCACAAAAAGAGA 4366  
Db 1721 GluArgSerGlnArgThrAspArgGlyThrGluProGlyProLysArgProSerHisArg 1740  
QY 4367 TGAGATTAACAGTGCAGGGTGGG-----GTCTGCATT 4399  
Db 1741 ProGlyProProValGlnGlnPheGlyThrSerAspLysAspSerAspLeuArgLeuVal 1760  
QY 4400 GGAGTGCCTTAAACAGAGAGAAATACTGAAGCAGACAGGGGAGGACACACAG 4459  
Db 1761 GlyAspSerLeuLysAlaGluLysGluLeuThrAlaSerValThrGluAlaIleProVal 1780  
QY 4460 ACCAGA-----CCAGGAGTCTCCAAACACAGAGTGGCAACAAA--- 4501  
Db 1781 SerArgAspTrpGluLeuLeuProSerAlaAlaSerAlaGluProGlnSerLysAsn 1800  
QY 4502 -----CCGAGCTGAGCATC-----AGACCTTGCCTCGA 4531  
Db 1801 LeuAspSerGlyHisCysValProGluProSerSerSerGlyGlnArgLeuTyProGlu 1820  
QY 4532 ATTGTCTTCCAGTATTCAGGTGCTCTTCTCTGCCCCCTTCCAGGGTATCTGGGTT 4591  
Db 1821 ValPheTyrglySerAlaGlyProSerSer-SerGlnIleSerGlyGlySerHisGlyLeu 1840  
QY 4592 G-----CCAGCTCG-----GGAGGCAACCATAGCCACAC---CACAG 4627  
Db 1840 uSerIleThrSerLysGlnTrpArgLeuArgProGlyThrProSerLeuHisProTyArg 1860  
QY 4628 GATTTCTGAAAGTTTACATGCTAGTATTTGGGTGTAGGTGGCAGCTCCCCAAG 4687  
Db 1860 G-----SerGlnProLeu 1864  
QY 4688 GCCCTGCCCC-----CCAGCCCCCAGCTCATGCTCTAAGTGTGTGTATTAATATT 4741  
Db 1864 uTyrgluProGlyProAlaProProSer-----AlaLeuLeuSerGln 1879  
QY 4742 TATTTATTGGAGATGTTTATTATTATAGATGATATTTATTGCAAGATTTCTATTCTGTAT 4801  
Db 1879 yValAlaLeuLysGlyGlnPhe-LeuAsp----- 1888  
QY 4802 TAACAAATAAATGCTGCCAGAACTAGTCTCTTTGCCAGC-----CTCACC 4852  
Db 1889 --PheSerThrMetGlnAlaThrGluLeuGly-LysLeuProAlaGlyGlyValLeuTy 1907  
QY 4853 CCTCTGTGTGCTCATCAGACTCTTGGCCCTTGGCTCCCACTCCCTGTCTGCTCCT--- 4907









RL Proc. Natl. Acad. Sci. U.S.A. 78:5334-5338(1981).  
RN [4]  
RP SEQUENCE OF 1-33 FROM N.A.  
RX MEDLINE=84297217; PubMed=6473103;  
RA Aho S., Tate V.E., Boedtker H.;  
RT "Location of the 11 bp exon in the chicken pro alpha 2(I) collagen  
gene.";  
RL Nucleic Acids Res. 12:6117-6125(1984).  
RN [5]  
RP SEQUENCE OF 1-79 FROM N.A.  
RX MEDLINE=88056316; PubMed=3678834;  
RA Finer M.H., Boedtker H., Doty P.;  
RT "Construction and characterization of cDNA clones encoding the 5' end  
of the chicken pro alpha 1(I) collagen mRNA.";  
RL Gene 56:71-78(1987).  
RN [6]  
RP SEQUENCE OF 78-92.  
RC TISSUE=Skin;  
RX MEDLINE=71115216; PubMed=5544653;  
RA Highberger J.H., Kang A.H., Gross J.;  
RT "Comparative studies on the amino acid sequence of the alpha 2-CB2  
peptides from chick and rat skin collagens.";  
RL Biochemistry 10:610-616(1971).  
RN [7]  
RP SEQUENCE OF 74-91; 263-448 AND 1088-1169 FROM N.A.  
RX MEDLINE=82058081; PubMed=6272119;  
RA Wozney J., Hanahan D., Tate V.E., Boedtker H., Doty P.;  
RT "Structure of the pro alpha 2 (I) collagen gene.";  
RL Nature 294:129-135(1981).  
RN [8]  
RP SEQUENCE OF 78-92.  
RC TISSUE=Skin;  
RX MEDLINE=70131186; PubMed=4313735;  
RA Kang A.H., Gross J.;  
RT "Amino acid sequence of cyanogen bromide peptides from the amino-  
terminal region of chick skicollagen.";  
RL Biochemistry 9:796-804(1970).  
RN [9]  
RP SEQUENCE OF 78-92 AND 415-448.  
RC TISSUE=Skin;  
RX MEDLINE=69285369; PubMed=5809220;  
RA Kang A.H., Igarashi S., Gross J.;  
RT "Characterization of the cyanogen bromide peptides from the alpha-2  
chain of chick skin collagen.";  
RL Biochemistry 8:3200-3204(1969).  
RN [10]  
RP SEQUENCE OF 78-92 AND 415-448.  
RC TISSUE=Bone;  
RX MEDLINE=69206882; PubMed=5785223;  
RA Lane J.M., Miller E.J.;  
RT "Isolation and characterization of the peptides derived from the  
alpha 2 chain of chick bone collagen after cyanogen bromide  
cleavage.";  
RL Biochemistry 8:2134-2139(1969).  
RN [11]  
RP SEQUENCE OF 566-587 FROM N.A.  
RX MEDLINE=79074929; PubMed=364479;  
RA Lehrach H., Frischauf A.-M., Hanahan D., Wozney J., Fuller F.,  
RA Kravchenkov R., Boedtker H., Doty P.;  
RT "Construction and characterization of a 2.5-kilobase procollagen  
clone.";  
RL Proc. Natl. Acad. Sci. U.S.A. 75:5417-5421(1978).  
RN [12]  
RP SEQUENCE OF 902-1362 FROM N.A.  
RX MEDLINE=81160715; PubMed=6927845;  
RA Fuller F., Boedtker H.;  
RT "Sequence determination and analysis of the 3' region of chicken pro-  
alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids  
including the carboxy-terminal propeptide sequences.";  
RL Biochemistry 20:996-1006(1981).  
RN [13]  
RP SEQUENCE OF 998-1169 AND 1234-1362 FROM N.A.  
RX MEDLINE=81264246; PubMed=6267043;

RA Dickson L.A., Ninomiya Y., Bernard M.P., Pesciotta D.M., Parsons J.,  
RA Green G., Eikenberry E.F., de Crombrughe B., Vogeli G., Pastan I.,  
RA Fietzek P.P., Olsen B.R.;  
RT "The exon/intron structure of the 3'-region of the pro alpha 2(I)  
collagen gene.";  
RL J. Biol. Chem. 256:8407-8415(1981).  
RN [14]  
RP SEQUENCE OF 932-954 AND 968-980 FROM N.A.  
RX MEDLINE=81064671; PubMed=6159982;  
RA Avvedimento V.E., Vogeli G., Yamada Y., Maizel J.V. Jr., Pastan I.,  
RA de Crombrughe B.;  
RT "Correlation between splicing sites within an intron and their  
sequence complementarity with U1 RNA.";  
RL Cell 21:689-696(1980).  
RN [15]  
RP SEQUENCE OF 126-161; 467-517 AND 926-954 FROM N.A.  
RX MEDLINE=81112157; PubMed=7460017;  
RA Yamada Y., Avvedimento V.E., Mudryj M., Ohkubo H., Vogeli G.,  
RA Irani M., Pastan I., de Crombrughe B.;  
RT "The collagen gene: evidence for its evolutionary assembly by  
amplification of a DNA segment containing an exon of 54 bp.";  
RL Cell 22:887-892(1980).  
RN CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
CC (FIBRILLAR FORMING COLLAGEN).  
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
CC HYDROXYAPATITE.  
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M25963; AAA69960.1; JOINED.  
DR EMBL; M25956; AAA69960.1; JOINED.  
DR EMBL; M25959; AAA69960.1; JOINED.  
DR EMBL; M25961; AAA69960.1; JOINED.  
DR EMBL; M25962; AAA69960.1; JOINED.  
DR EMBL; M25965; AAA69961.1; JOINED.  
DR EMBL; M25964; AAA69961.1; JOINED.  
DR EMBL; M25984; AAA69962.1; JOINED.  
DR EMBL; M25957; AAA69962.1; JOINED.  
DR EMBL; M25966; AAA69962.1; JOINED.  
DR EMBL; M25967; AAA69962.1; JOINED.  
DR EMBL; M25969; AAA69962.1; JOINED.  
DR EMBL; M25970; AAA69962.1; JOINED.  
DR EMBL; M25971; AAA69962.1; JOINED.  
DR EMBL; M25972; AAA69962.1; JOINED.  
DR EMBL; M25973; AAA69962.1; JOINED.  
DR EMBL; M25974; AAA69962.1; JOINED.  
DR EMBL; M25976; AAA69962.1; JOINED.  
DR EMBL; M25977; AAA69962.1; JOINED.  
DR EMBL; M25978; AAA69962.1; JOINED.  
DR EMBL; M25979; AAA69962.1; JOINED.  
DR EMBL; M25980; AAA69962.1; JOINED.  
DR EMBL; M25981; AAA69962.1; JOINED.  
DR EMBL; M25982; AAA69962.1; JOINED.  
DR EMBL; M25983; AAA69962.1; JOINED.  
DR EMBL; J00826; AAA51611.1; JOINED.  
DR EMBL; J00821; AAA51611.1; JOINED.  
DR EMBL; J00792; AAA51611.1; JOINED.  
DR EMBL; J00830; AAA51613.1; JOINED.  
DR EMBL; J00829; AAA51613.1; JOINED.  
DR EMBL; J00837; AAA51614.1; JOINED.  
DR EMBL; J00812; AAA51615.1; JOINED.  
DR EMBL; J00811; AAA51615.1; JOINED.  
DR EMBL; J00814; AAA51615.1; JOINED.

DR EMBL; J00815; AAAS1615.1; JOINED.  
 DR EMBL; X02657; CAA26493.1; -  
 DR EMBL; K00794; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; V00390; CAA23688.1; -  
 DR EMBL; X00760; CAA25330.1; -  
 DR EMBL; M17608; AAA48673.1; -  
 DR EMBL; M10581; AAA48637.1; -  
 DR EMBL; M10540; AAA48638.1; -  
 DR EMBL; J00828; AAAS1612.1; -  
 DR EMBL; J00827; AAAS1612.1; JOINED.  
 DR EMBL; J00831; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; J00832; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; J00833; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; J00822; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A02868; CGCH28.  
 DR PIR; S07354; S07354.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001670; Fe-ADH.  
 DR InterPro; IPR000885; Fib collagen C.  
 DR InterPro; IPR001431; Peptidase M16.  
 DR InterPro; IPR000244; Ribosomal\_L9.  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR InterPro; IPR000360; Transketolase.

Alignment Scores:  
 Pred. No.: 4.84e-07 Length: 1362  
 Score: 306.50 Matches: 354  
 Percent Similarity: 30.58% Conservative: 86  
 Best Local Similarity: 24.60% Mismatches: 520  
 Query Match: 3.28% Indels: 480  
 DB: 1 Gaps: 72

US-09-931-704-3 (1-5087) x CA21\_CHICK (1-1362)

Qy 4305 GGGGATAGGAGGAGCAACCCATCCAGAGAGTGGGGAGGCCAAAGGCTTCACCCAG 4246  
 Db 37 GlyAspLysGlyPro-----GlnGlyGluArgGly--ProProGly---ProProGly 52  
 Qy 4245 AAAGCATGTGTGGTGGGCGCCCTCAACCCCTGCGCTCCAGATGTGCCA---CCTGAGG 4189  
 Db 52 LysAspGlyGluAspGlyPro-----ProGlyProProGlyProProGlyProProG 70  
 Qy 4188 GCGTGGTGGGCACTGG----- 4172  
 Db 70 LysGlyGlyAsnPheAlaGlnTyrAspProSerLysAlaAlaAspPheGlyProG 90  
 Qy 4171 -----CCAA 4168  
 Db 90 LysProMetGlyLeuMetGlyProArgGlyProProGlyProGlyAlaSerGlyProProGlyProP 110  
 Qy 4167 GTGGTAGGCAAGAGCAGGAGCCATGAGCGCTCTGCACCAACCTGAACCACTTCACAC 4108  
 Db 110 roGlyPheGlnGlyValProGlyGlu----- 118  
 Qy 4107 TCCCTCGAGCATGACTTCTGTAGAAACAGGACAGGAGCGGCTTACTGTACTT----- 4054  
 Db 119 -----ProGlyGluProGlyGln-ThrGlyProGlnGlyProArgGly 132  
 Qy 4053 CCTCCCACTCGGTAGACCTTTGGAGGTGGGGAGGAGACAGGCGTGTATCCATCACAC 3994  
 Db 133 ProPro-----GlyProProGlyLysAlaGlyGluAspGlyHisProGlyLysPro 149  
 Qy 3993 GCCAGAGCGGTCCAGGAAGGCGCAGAGGCTCACAGCTTCTGTCT-----CCTGGCTCA 3940  
 Db 150 GlyArgProGlyGluArgGlyValAlaGlyProGlnGlyAlaArgGlyPheProGlyThr 169  
 Qy 3939 ACAGGT---GTTGGCATACAGGCTGGCTCTCACAAGAGTGAGGAG-CAGGGTTGAAGGGG 3884  
 Db 170 ProGlyProProGlyPheLysGlyLysGlyHisAsnGlyLeuAspGlyLeuThrGly 189  
 Qy 3883 GAGCCAGAGGAGAGGTCAAGAGTCAGAGCCATGAGCCAGGCGCCAGGTCAGGCGTCACTGC 3824  
 Db 190 GlnProGlyAlaProGlyThrLysGlyGluProGlyAlaProGlyGluAsnGly-ThrPr 209

Qy 3823 AGCTGTGGAGGTGTCATCTTCTTTGAGCCGGTTGAAGTCTTGGCCGAGCCGACAG 3764  
 Db 209 oGlyGlnProGlyAla-----ArgGlyLeu-ProGlyGluArg 222  
 Qy 3763 CCAGGTCTGAGCTCTCTTACAGCCAGAGTCTCTCATCTTCTGGAGGA----- 3714  
 Db 222 LysGlnGlyAlaProGlyProAlaGlyAlaArgGlySerAspGlySerAlaGlyPro 242  
 Qy 3713 -----AGTCACTGTGGGAGGCGCAGGAGTCAAGTGGGTTTCAAGTCCAGGCGG 3662  
 Db 242 hrGlyProAla\*\*\*\*\* 262  
 Qy 3661 CTGGGCGAGTGGT-----AGCCAGAGCTGCCATGACGCCGCAATGCTGCCAGCAG 3608  
 Db 262 \*GlyGlnGlyProAlaGlyAsnGlyPro-----ThrGlyProAlaG 278  
 Qy 3607 GCTCTGAGGCTGTGTC-----AGAAGTGGGCGAGGCTGGCGCCAGCTCAGC 3560  
 Db 278 LysProArgGlyGluLeuGlyLeuProGlySerSerGlyProVal----- 292  
 Qy 3559 AGTGGCAGCTGACGGTTGAGGCGCAGCAAGTAACACAGAAGGTGGTGTAGGCTCGTA 3500  
 Db 293 -----GlyProProGlyAsnProGlyAlaAsnGlyLeuProGlyAla 306  
 Qy 3499 GTTCTGGGTGAGCGCAGTTTGTTCATTGAGGCTTTCGCCACACCTCCAAAGTCAACAGTGGC 3440  
 Db 306 LysGlyAlaAlaGlyLeuPro-----GlyValAlaGlyAlaProGlyLeuProGlyP 324  
 Qy 3439 CCTGGGAGAGTCTCTGCCCGGAGGGGAGGTTGAAGTCTGGCTGCT---TGAAAGG 3383  
 Db 324 roArgGlyIleProGlyProProGlyProAlaGlyProSerGlyAlaArgGlyLeuValG 344  
 Qy 3382 GGGGCCAGGTAGTTCAGCTGTGAAAGGAGGAGGTGATGGGAGGAGGAGGAGGAGG 3323  
 Db 344 LysGlyProGlyProAlaGlyAla-LysGlyGlySerGlyAsnLysGly-GluProGlyAla 363  
 Qy 3322 CGCTGGCTCACCAAGGAGATACCTGCTCCCAAGTTTCTATTTTGTGTGCCCT--- 3265  
 Db 363 aAlaGlyProPro-----GlyProPr 370  
 Qy 3264 -----GA 3263  
 Db 370 oGlyProSerGlyGluGlyLysArgGlySerAsnGlyGluProGlySerAlaGlyPr 390  
 Qy 3262 CACTGGCCCTGTCTCCATGCTAGGCTTC---TTTGTTCATGGCTCCCT---GAGTTTTT 3209  
 Db 390 oProGlyProAlaGlyLeuArgGlyGluProGlySerArgGlyLeuProGlyAlaAspG 410  
 Qy 3208 CCAATAAGGATATCTTTGTATGACAGTGACAGTTGAAGCAGTGGC-----GGTGGTGA 3155  
 Db 410 YArgAlaGlyValMetGlyProAlaGlyAsnArgGlyAlaSerGlyProValGlyAlaL 430  
 Qy 3154 GCTCTCCATGAGGAGGCGTAGAAAAGAGGAG----- 3121  
 Db 430 sGlyProAsnGlyAspAlaGlyArgProGlyGluProGlyLeuMetGlyProArg\*\*\*\*\* 450  
 Qy 3120 -----GAGGAGAGCGCACAGACGCTCAGGAAGGAGCGTCTGATTGGCTGGGTGGG 3071  
 Db 450 \*\*\*\*\*GlyPheProG 470  
 Qy 3070 CAAACCTTGGAGAGTCCCGCTGTGGG---GCAGGACAGAGAGAGAGAG----- 3023  
 Db 470 YAlaAsp-GlyArgValGlyProLleGlyProAlaGlyAsnArgGly-GluProGlyAsn 489  
 Qy 3022 -----GAGCCTGGAGGAGCTCGAGAGTAGAA 2996  
 Db 490 lLeGlyPheProGlyProLysGlyProThrGlyGluProGlyLysProGlyGlu---Lys 508  
 Qy 2995 GGTGGCCAGGAGAGTGCAGAGACCAAGAGAGGAGTTTCAGAGAGGAGCAGCGCTCAG 2936  
 Db 509 GlyAsnValGlyLeuAlaGlyProArgGlyAlaProGlyProGlyGlyAsnGlyAla 528







US-09-931-704-3 (1-5087) x CAL1\_MOUSE (1-1453)

QY	5059	GGGAAACCTGGGGTGGACAGAAAGCTGTGTGGGAGGACTGAGGATCAAGGGGTGTGGTGT	5000
DB	192	GlyProProGly-----GluProGlyGluProGly	201
QY	4999	GGCAGCAGCGGTGTAAGTCTGAGGTGCGCCATCGCCCGTGGGGAGGAGAGG	4940
DB	202	GlySer-----GlyProMetGlyPro---ArgGlyProProGly	213
QY	4939	CCTCTTCCAGAGCTCTGTGCAGCTCCACAGAGCAGCAGGAGTGGAGCCAGGGT	4880
DB	214	Pro-----ProGlyLysAsnGlyAspAspGlyGluAla	224
QY	4879	GGCAAGAGCTCTGATGAGCACCAGGAGGGGTGAGGTGGGCAAGAGAGACTAAGTCTTGGGG	4820
DB	225	GlyLysProGlyArgProGlyGluArgGlyProProGlyProGlnGly---AlaArgGly	243
QY	4819	CAAGCATTTATTGTTTAATACAAGATAGAAATCTGCAATAAATATCATCTAATAAT	4760
DB	244	LeuPro-----	245
QY	4759	AACATCTCCAAATAATAATTAATATACACACACTTAGAGTCATGAGTGGTGGGCT	4700
DB	246	-----GlyThrAlaGlyLeuProGlyMetLysGlyHisArg	257
QY	4699	GGGGGCGAGGCCTT-----GGGGAGCTGCCACCTACACCCCAAAATGCTAC	4652
DB	258	GlyPheSerGlyLeuAspGlyAlaLysGlyAspAlaGlyProAlaGlyProlys-----	275
QY	4651	TGCATTGTAACTTTCAGGAATCTGTGGTGTGGCTATGTTGCTCCCTCCAGCCTGGC	4592
DB	276	-----GlyGluProGlySerProGly	282
QY	4591	AACCCACAGATACCCTGGGAAGGGGCGAGAGAGGACCGCTAATACTGGAAGACAAT	4532
DB	283	GluAsnGlyAlaProGlyGlnMetGlyProArgGly-----LeuProGlyGluArg	299
QY	4531	TCGAGGCAAGTCTCT-----GATGCTCAGCTCGGGTTT	4499
DB	300	GlyArgProGlyProProGlyThrAlaGlyAlaArgGlyAsnAspGlyAlaValGlyAla	319
QY	4498	TGTTTGCCACTCTGTGCTTTGGAGACTCTCGGTCTGGTCTGGTCTGCTCCCTGCCCTGT	4439
DB	320	AlaGlyPro-----ProGly	324
QY	4438	GCTTTCAGTATTTCTCTCTGTTTATAGGCGACCTCCAATGCGACACCCCAACCTGCA	4379
DB	325	-----ProThrGlyProThrGly-ProPro-----	332
QY	4378	CTGTTAATCTCATCTCTTTTGTGTAGAAATGCCACCATGTTGTTCTCTGAAATGGATG	4319
DB	332	-----	332
QY	4318	TATAGAGTGAATTTGGGATAAGGAGCAACACCCATCAAGGAGAGTGGGCGAGCAAG	4259
DB	333	-----GlyPheProGlyAlaValGlyAlaLysGlyGluAlaGlyProGlnGly	348
QY	4258	GGCTTCAACCCAGAAAGCATGTGTGT-----GGTGGCCCTCAACCCCTGC	4211
DB	348	AlaArgGlySerGluGlyProGlnGlyValArgGlyGluProGlyProGlyProAla	368
QY	4210	CCTCCAGATGCGCCACTGAGGGCTGGTGGGC-----ACTGG	4172
DB	368	a-----GlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGly	381
QY	4171	CCAAGTGTAGGCAAGAGCAGGAG-----	4147
DB	381	YGlnProGlyAlaLysGlyAlaAsnGlyAlaProGlyIleAlaGlyAlaProGlyPhePr	401
QY	4146	---GCCATGAGCGCTCTGCACCAACCTGAACCACTTCACTCTCCCTCGAGCATGACTT	4091



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1] NCBI\_TaxID=9606;  
RP SEQUENCE FROM N.A. (ISOFORM SMRT).  
RC MEDLINE=99178941; PubMed=10077563;  
RX ORDLINE=99178941; PubMed=10077563;  
RA Ordlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;  
RT "Unique forms of human and mouse nuclear receptor corepressor SMRT.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644 (1999).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM SMRT).  
RC TISSUE=Cervical adenocarcinoma;  
RX MEDLINE=9919215; PubMed=10097068;  
RA Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;  
RT "SMRT, a silencing mediator for retinoid and thyroid hormone  
RT receptors-extended isoform that is more related to the nuclear  
RT receptor corepressor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524 (1999).  
RN [3]  
RP SEQUENCE OF 1023-2517 FROM N.A.  
RC TISSUE=Cervical adenocarcinoma;  
RX MEDLINE=96008552; PubMed=7566127;  
RA Chen J.D., Evans R.M.;  
RT "A transcriptional co-repressor that interacts with nuclear hormone  
RT receptors.";  
RL Nature 377:454-457 (1995).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM TRAC-1).  
RC TISSUE=Fetal liver;  
RX MEDLINE=96408715; PubMed=8813722;  
RA Sande S., Privalsky M.L.;  
RT "Identification of TRACs (T3 receptor-associating cofactors), a family  
RT of cofactors that associate with, and modulate the activity of,  
RT nuclear hormone receptors.";  
RL Mol. Endocrinol. 10:813-825 (1996).  
RN [5]  
RP SEQUENCE OF 428-613 FROM N.A.  
RC TISSUE=Brain cortex;  
RX MEDLINE=9736492; PubMed=9225980;  
RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,  
RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;  
RT "CDNAs with long CAG trinucleotide repeats from human brain.";  
RL Hum. Genet. 100:114-122 (1997).  
CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIB.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SMRT/TRAC-2 (SHOWN HERE) AND  
CC THE C-TERMINAL RECEPTOR-INTERACTING DOMAIN AND ACTS AS AN  
CC ANTIREPRESSOR.  
CC -!- TISSUE SPECIFICITY: UBIQUITOUS. HIGH LEVELS OF EXPRESSION ARE  
CC DETECTED IN LUNG, SPLEEN AND BRAIN.  
CC -!- INDUCTION: REGULATED DURING CELL CYCLE PROGRESSION.  
CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
CC DOMAINS (ID1 AND ID2).  
CC -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND  
CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES  
CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
CC SPECIFICITY.  
CC -!- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 CORNR BOXES.  
CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS

CC FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF113003; AAD20946.1; -;  
DR EMBL; AF125672; AAC22973.1; -;  
DR EMBL; U37146; AAC50236.1; -;  
DR EMBL; S83390; AAB50847.1; -;  
DR EMBL; U80750; AAB91446.1; -;  
DR TRANSFAC; T04689; -;  
DR Genew; HGNC:7673; NCOR2.  
DR MIM; 600848; -;  
DR InterPro; IPR001005; Myb DNA binding.  
DR Pfam; PF00249; myb DNA-binding; 2.  
DR SMART; SM00395; SANT; 2.  
DR PROSITE; PS00909; MYB\_3; 1.  
KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
KW Coiled coil; Alternative splicing.  
FT DOMAIN 174 215  
FT DOMAIN 254 312  
FT DNA\_BIND 429 474  
FT DNA\_BIND 513 557  
FT DOMAIN 522 561  
FT DOMAIN 778 820  
FT DOMAIN 2139 2143  
FT DOMAIN 2342 2346  
FT DOMAIN 494 510  
FT DOMAIN 682 685  
FT DOMAIN 994 1002  
FT DOMAIN 1384 1389  
FT DOMAIN 1842 1846  
FT DOMAIN 2479 2482  
FT VARSPIC 1 1702  
FT VARSPIC 2353 2398  
FT CONFLICT 295 295  
FT CONFLICT 309 309  
FT CONFLICT 352 352  
FT CONFLICT 365 365  
FT CONFLICT 612 613  
FT CONFLICT 711 711  
FT CONFLICT 724 740  
FT CONFLICT 787 796  
FT CONFLICT 804 804  
FT CONFLICT 814 814  
FT CONFLICT 817 817  
FT CONFLICT 889 889  
FT CONFLICT 1023 1030  
FT CONFLICT 1034 1034  
FT CONFLICT 1894 1894  
FT CONFLICT 2494 2494  
SQ SEQUENCE 2517 AA; 274031 MW; F5805C01761258C0 CRC64;  
Alignment Scores:  
Pred. No.: 4.62e-07 Length: 2517  
Score: 306.50 Matches: 420  
Percent Similarity: 30.52% Conservative: 148  
Best Local Similarity: 22.57% Mismatches: 600  
Query Match: 3.25% Indels: 693  
DB: 1 Gaps: 99  
US-09-931-704-3 (1-5087) x NCR2\_HUMAN (1-2517)  
QY 37 GCTTCGCGGAGCGCGGCTC-----GCCCTCCCA 66  
Db 955 AlaserProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaIlePro 974







RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;  
 RT "Organization of the exons coding for pro alpha 1(II) collagen N-  
 RT propeptide confirms a distinct evolutionary history of this domain of  
 RT the fibrillar collagen genes.";  
 RL Genomics 4:438-441(1989).  
 RN [10]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=91184577; PubMed=2010058;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in collagen genes: causes of rare and some common diseases  
 RT in humans.";  
 RL FASEB J. 5:2052-2060(1991).  
 RN [11]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97255959; PubMed=9101290;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 RN [12]  
 RP VARIANT SER-1074.  
 RX MEDLINE=97255959; PubMed=2572591;  
 RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,  
 RA Hollister D.W.;  
 RT "Glycine to serine substitution in the triple helical domain of pro-  
 RT alpha 1 (II) collagen results in a lethal perinatal form of short-  
 RT limbed dwarfism.";  
 RL J. Biol. Chem. 264:18265-18267(1989).  
 RN [13]  
 RP VARIANT SEDC 1095-GLY--TYR-1330 DEL.  
 RX MEDLINE=89286907; PubMed=2543071;  
 RA Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;  
 RT "Identification of the molecular defect in a family with  
 RT spondyloepiphyseal dysplasia.";  
 RL Science 244:978-980(1989).  
 RN [14]  
 RP VARIANT OSTEOARTHRITIS CVS-650.  
 RX MEDLINE=90370826; PubMed=1975693;  
 RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;  
 RT "Single base mutation in the type II procollagen gene (COL2A1) as a  
 RT cause of primary osteoarthritis associated with a mild  
 RT chondrodysplasia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).  
 RN [15]  
 RP VARIANT OI-IV VAL-717.  
 RX MEDLINE=91291136; PubMed=2064612;  
 RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;  
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to  
 RT valine substitution in osteogenesis imperfecta type IV. Detection of  
 RT the mutation and prenatal diagnosis by a chemical cleavage method.";  
 RL Biochem. J. 276:765-770(1991).  
 RN [16]  
 RP VARIANT OSTEOARTHRITIS CVS-650.  
 RX MEDLINE=91086471; PubMed=1985108;  
 RA Eyre D.R., Weis M.A., Moskowitz R.W.;  
 RT "Cartilage expression of a type II collagen mutation in an inherited  
 RT form of osteoarthritis associated with a mild chondrodysplasia.";  
 RL J. Clin. Invest. 87:357-361(1991).  
 RN [17]  
 RP VARIANT HYPOCHONDROGENESIS GLU-984.  
 RX MEDLINE=93054548; PubMed=1429602;  
 RA Bogart R., Tiller G.E., Weis M.A., Gruber H.E., Rimoin D.L.,  
 RA Cohn D.H., Eyre D.R.;  
 RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha  
 RT 1(II) chain produces hypochondrogenesis.";  
 RL J. Biol. Chem. 267:22522-22526(1992).  
 RN [18]  
 RP VARIANT HYPOCHONDROGENESIS SER-705.  
 RX MEDLINE=92262484; PubMed=1374906;  
 RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,  
 RA Ramirez F., Vitale E., Lee B.;  
 RT "Characterization of a type II collagen gene (COL2A1) mutation

RT identified in cultured chondrocytes from human hypochondrogenesis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).  
 RN [19]  
 RP VARIANT WS-II ASP-198.  
 RX MEDLINE=93304428; PubMed=8317498;  
 RA Koerkoe J., Ritvaniemi P., Haataja L., Kaaerlaeinen H.,  
 RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;  
 RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate  
 RT for glycine alpha 1-67 and that causes cataracts and retinal  
 RT detachment: evidence for molecular heterogeneity in the Wagner  
 RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";  
 RL Am. J. Hum. Genet. 53:55-61(1993).  
 RN [20]  
 RP VARIANT SEMD CVS-840.  
 RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,  
 RA Eyre D.R.;  
 RT "A dominant mutation in the type II collagen gene (COL2A1) produces  
 RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";  
 RL Am. J. Hum. Genet. 53:A209-A209(1993).  
 RN [21]  
 RP VARIANT OSTEOARTHRITIS CVS-650.  
 RX MEDLINE=93282819; PubMed=8507190;  
 RA Holderbaum D., Malemud C.J., Moskowitz R.W., Haqqi T.M.;  
 RT "Human cartilage from late stage familial osteoarthritis transcribes  
 RT type II collagen mRNA encoding a cysteine in position 519.";  
 RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).  
 RN [22]  
 RP VARIANT SEMD ARG-285.  
 RX MEDLINE=93252400; PubMed=8486375;  
 RA Viikula M., Ritvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,  
 RA Peltonen L.;  
 RT "A mutation in the amino-terminal end of the triple helix of type II  
 RT collagen causing severe osteochondrodysplasia.";  
 RL Genomics 16:282-285(1993).  
 RN [23]  
 RP VARIANT SEDC CVS-206.  
 RX MEDLINE=94063862; PubMed=8244341;  
 RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,  
 RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;  
 RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a  
 RT family with an Arg75-->Cys mutation in the procollagen type II gene  
 RT (COL2A1).";  
 RL Hum. Genet. 92:499-505(1993).  
 RN [24]  
 RP VARIANT SEDC CVS-920.  
 RX MEDLINE=93315508; PubMed=8325895;  
 RA Chan D., Taylor T.K.F., Cole W.G.;  
 RT "Characterization of an arginine 789 to cysteine substitution in  
 RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal  
 RT dysplasia.";  
 RL J. Biol. Chem. 268:15238-15245(1993).  
 RN [25]  
 RP VARIANT SEDC SER-1128.  
 RX MEDLINE=93140139; PubMed=8423604;  
 RA Cole W.G., Hall R.K., Rogers J.G.;  
 RT "The clinical features of spondyloepiphyseal dysplasia congenita  
 RT resulting from the substitution of glycine 997 by serine in the alpha  
 RT 1(II) chain of type II collagen.";  
 RL J. Med. Genet. 30:27-35(1993).  
 RN [26]  
 RP Alignment Scores:  
 Pred. No.: 5.62e-07 Length: 1418  
 Score: 305.00 Matches: 302  
 Percent Similarity: 21.10% Conservative: 64  
 Best Local Similarity: 35.66% Mismatches: 376  
 Query Match: 3.27% Indels: 437  
 DB: 1 Gaps: 68  
 US-09-931-704-3 (1-5087) x CA12\_HUMAN (1-1418)  
 QY 4305 GGGGTAAGGAGGACCAACACCCATCCAGAGAGTGGGCGAGCCAAAGGCGCTTCAACCCAG 4246  
 Db 60 GlyGluGlnGlyProArgGlyAspArgGlyAspLysGlyGlyGlyAlaProGlyPro 79



QY	4245	AAAGGCGTGTGGT-----GGTGGCCCTCAACCCCTGC-CTCCAGAT	4202
Db	80	ArgGlyArgAspGlyGluProGlyThrLeuGlyAsnProGlyProProGlyProProGly	99
QY	4201	GTCCCA---CCTGAGGGCTGGGTGGGCACCTGGCCAAAGTGGTAGCAAGAGCAGGAGGC	4145
Db	100	ProProGlyProProGlyLeuGlyGlyAsnPheAlaAla-----GlnMetAlaGlyGly	117
QY	4144	CATGAGGCGCTCTGC-----ACCAACTGAACACCTTCACACTCCCTCGAGCATGACTT	4091
Db	118	PheAspGluIysAlaGlyAlaGlnLeuGlyValMetGlnGlyProMetGlyProMet	137
QY	4091	-----	4091
Db	138	GlyProArgGlyProProGlyProAlaGlyAlaProGlyProGlnGlyPheGlnGlyAsn	157
QY	4090	-----CCTGTAAACAGCAGCAGCAGGCGCTACTGTACTCTCTCTCC	4049
Db	158	ProGlyGluProGlyGluProGlyValSerGlyPro-MetGlyProArgGlyProProGly	177
QY	4048	CCAGCTCGGTAGACTTTGGAGGTGGGAG-----GAGCAGGGCTGATCGCATCACA	3995
Db	177	yProProGlyIysProGlyAspAspGlyGluAlaGlyIysProGlyIysAlaGlyGluAr	197
QY	3994	CGCCAGCCGCTCCAGAAAGGCGCAGAGGCTCACAGCTTCTGTCTCTCTCTCTCAACAGG	3935
Db	197	gGlyProProGlyProGlnGlyAlaArgGly-----PheProGlyThrProGly	213
QY	3934	TGTT---GGCATACAGGGCTGGCTCTCAAAAGTGGAG-CAGGTTTGAAGGGGAGCG	3879
Db	213	yLeuProGlyValIysGlyHisArgGlyTyPProGlyLeuAspGlyAlaIysGlyGluAl	233
QY	3878	AAGAGAGAGGTCAGAAAGTCAGAAAGCAGCCAGCCAGGTCAGGCTGACTGCAGCTG	3819
Db	233	aglyAlaProGlyValIysGlyIuserGlySer--ProGly-----	246
QY	3818	CTGAGGCTGCATCTTCTCTTGAGCGGTTGAAGTCTCTGCGCGAGCGCCACAGCCAGG	3759
Db	247	---GluAsnGlySerProGlyProMetGlyProArgGlyLeuProGlyGluArgGlyArg	265
QY	3758	TCTCAGCTCTTCAGCAGCCAGAAAGTCGTCCATCTCTGAGAGGAAGTCATGTGGGCGAG	3699
Db	266	ThrGlyProAlaGlyAlaAlaGlyAlaArgGlyAsnAspGly-----	279
QY	3698	GGCCAGGAGTCCAAGTGGTTAGTCCCGCAGCAGCGGCTGGGCGAGTGGTAGCCAGAG	3639
Db	280	-----GlnProGlyProAlaGlyProProGly	288
QY	3638	CTGCCATGACCCCGCAATGCTGCTCCAGCAGCCCTCGAGGCTGGTGCAGAAAGTGGGCCA	3579
Db	289	-----ProValGlyProAlaGly-----GlyPro	296
QY	3578	GGCTCGCGCGAGCTCAGCAGTGGCAGCTGACGGTTGAGGCCACGCCAAGTAACACAGAA	3519
Db	297	GlyPheProGlyAla-----ProGlyAlaIysGlyGluAlaGlyProThr---	311
QY	3518	GGTGGCTGAGCCCTGAGTCTCTGGTCCAGCCGAGTTTGTATGAGGCTTCGCCACA	3459
Db	312	-----GlyAlaArgGly	315
QY	3458	CCTCCAAGTCAACGTGGCCCTGGGCAGAGTCTCTGCCCGCCAGCGGGGAGGTTGAAGT	3399
Db	316	ProGluGlyAlaGlnGlyProArgGlyGluProGlyThrProGlySerProGlyProAla	335
QY	3398	CTGGCTCGTTGAAGGGGGGGCCAGGTAGTTACGTGCTGTGAAGAGGAGGGGTATGGGA	3339
Db	336	GlyAla-----SerGlyAsnProGlyThrAspGlyIlePro-----GlyAlaIysGly	351
QY	3338	AGGAGGAGGCGAGCCGTGCTCACACAGGAGATACCTGTCTCCCAAAAGTTCTAT	3279
Db	352	SerAlaGlyAlaProGlyIleAlaGlyAlaProGly-PheProGlyProArg-----	368





QY 2915 CAGCAAGGAGAAATGAGCCTTCAAACAACACTGTCGGGCTTGGCAAGAGGTGCTGACGACC 2856  
 Db 539 -----ArgGlyGlyAlaGlyProPr 545  
 QY 2855 TTGAGAGAACCAATGCTGGGCTTGGAGGAGGAGGCTTAAAGAGTGAAGTGG 2796  
 Db 545 o-----GlyPro-----GluGlyGlyLysGlyAlaAlaG 555  
 QY 2795 GTGAGAGAGAGTGGAGCAGAGGTGTAGACCCTCTTGGAGAAAATTTGGCAGGAAA 2736  
 Db 555 yProProGlyProProGlySerAlaGlyThrProGlyLeu-----GlnGlyLe 571  
 QY 2735 GCCAGGAGAGAAA-----AAGGTCGGGAACT 2709  
 Db 571 tProGlyGluArgGlyGlyProGlyGlyProGlyProLysGlyAspLysGlyGlyProG 591  
 QY 2708 AGAGAGGCGTGTGAGTGAACAAAGTTTCTGCACAAAGAAACAGAGAGATGAGAG 2649  
 Db 591 ySerSerGlyValAspGlyAlaProGly-----LysAspGlyProArgGlyProTh 608  
 QY 2648 TGGCCCATTTGGAACAGAGCCCATTTGCGCTGCACACACAGCCAGCCACAGCTGAGGG 2589  
 Db 608 rGlyProIleGlyProProGly-----ProAlaGlyGlnProGlyAspLysGlyGly 626  
 QY 2588 GCTGGGAGAGACGGGTAC---TCACTAACACTCCGAAGAAGGGTTTCAGCTGATAGGAG 2532  
 Db 626 x---GlyAlaProGlyValProGlyIleAlaGlyProArgGlyGlyProGlyGluArg 645  
 QY 2531 CAGGGTT-----GCAGAAAGAAAGTATAGGAGAGAGGAGG 2502  
 Db 645 yGluGlnGlyProProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGly----- 663  
 QY 2501 TCTCATGAGTAGGACAGAGAGAGGCTATGAACAGACAGCATGTTATAGGAGAGGAGG 2442  
 Db 664 -----GluProGly-AlaLysGlyGluArgGlyAlaProGlyGluLysGlyG 679  
 QY 2441 AGCCCTTTCCAAACCTCGGGGACGCTGGGCGCCAGGGAAGGATGAGGAGATGTA 2382  
 Db 679 lu-----GlyGlyProProGlyAlaAlaGlyProAlaGlyGlySerGlyP 694  
 QY 2381 TAACTGGGAAGAAGGAG-----AAGCAAGGGAGAGG-----CTG 2344  
 Db 694 roAlaGlyProProGlyProGlnGlyValLysGlyGluArgGlySerProGlyGlyProG 714  
 QY 2343 GGATA-----GGCAAGGGCCAGGGGAGGTTACACGCTGACAAGCACA 2299  
 Db 714 lyAlaAlaGlyPheProGlyGlyArgGlyProProGlyPro----- 727  
 QY 2298 GAGACAACTAATTTCTCGGTATCTCGGGCTGGAGCCAGCTCCCGCCACCTTCT 2239  
 Db 728 -----ProGlySerAsnGlyAsn-ProGlyProProGly 738  
 QY 2238 TCCCTTTTCTTCTCCGCTCCAGATAGGTCTGGGATCTGGAGAGGAGACAGGAG 2179  
 Db 739 SerSerGlyAlaPro-----GlyLys 745  
 QY 2178 AGGGGCGCGGT---CCCCATGTTGCTCGGCTCTCCACCCCTCTCTCTCTCTCTG 2122  
 Db 746 AspGlyProProGlyProProGlySerAsnGlyAlaProGly---SerProGlyLys 764  
 QY 2121 GGTCTCTGNACACCTCTGGACCAACCCCAATCTTACAGCATAACTCCATCCACACC 2062  
 Db 765 GlyProLysGlyAspSerGlyPro----- 772  
 QY 2061 CTAACCCCTCGCCACTACTGCTGGGTTCCTACAGTCCCTTCCCTCCATGACCCCT 2002  
 Db 773 -----ProGlyGluArgGlyAlaProGlyProGlnGlyProPro 785  
 QY 2001 AGGAGCCACAGTAGAGTGGAGAGGAGCAGGAGGAGGCC----- 1959  
 Db 786 -GlyAlaProGlyProLeuGlyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyPr 805

QY 1958 -----CCTGGGTCCAGACGACGA---GGGCCCCAG---GTCTGGGGGCGAGAGCG 1915  
 Db 805 oProGlyMetProGlyAlaArgGlySerProGlyProGlnGlyLysGlyGlyGluAsn-- 824  
 QY 1914 AGCTGAACCCAGGATGGAGGTGGGGCCCTCTGTCTCCCAACCCCGCCCGCGCA 1855  
 Db 825 -----GlyLysProGlyPro----- 829  
 QY 1854 AGAGACACCAACACAGCTGTAAACAGAGAGACTTCCCGGCCCCAGGGGAGGGGG 1795  
 Db 830 ---SerGlyGlnAsnGly---GluArgGlyPro---ProGlyProGlnGlyLeuProG 846  
 QY 1794 C-----GCTGGGCTGGCAGCCAGTGGGGCAATGGCCCAAGGGGTGGAGGCTTCC 1741  
 Db 846 yLeuAlaGlyThrAlaGlyGlu-ProGlyArgAspGlyAsnProGlySerAspGlyLeuP 866  
 QY 1740 CATATCCCTAGTCTCTGGGCCAGGATGGGAGTGGGGGCCACCTTAAGAGAGGAGGG 1681  
 Db 866 ro-----GlyArgAspGlyAla-----ProGlyAlaL 875  
 QY 1680 CAGGAGATAGACAGACAGGGGACTGTGGAGCCCAAGAGCCCTCTCTCACCCCATCA 1621  
 Db 875 ysGlyAspArgGlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisPro--- 893  
 QY 1620 TTTACCAAGAACCTCCA-----CGGTTAGGACTGTCTTCCCAACTCTCTCAC 1573  
 Db 894 -----GlyProProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGlu 911  
 QY 1572 TCCTCCCAACTCCAGATTCCTACCTGGATCTACATAGCTCCAGCAAGAGCTCGGG 1513  
 Db 911 hrGlyPro-----AlaGlyProSerGlyAlaPro- 920  
 QY 1512 AGTTGGTCTCAGGTAGCGGTGAGGTCTAGTTTCTGATGAGGGGCCAGGCCCT 1453  
 Db 921 -----GlyProAlaGlySer-----ArgGly-ProProGlyPro 931  
 QY 1452 GGGTCCCTGTGCGATTGAGAGCTGGCACTGCAGGAGGTGCAGAGCACCGTGACAGG 1393  
 Db 932 GlnGlyProArgGlyAspLysGlyGluThrGlyGluArgGlyAlaMetGlyLysGly 951  
 QY 1392 CACGCTAACATCCCT-----CAGCAG 1372  
 Db 952 HisArgGlyPheProGlyAsnProGlyAlaProGlySerProGlyProAlaGlyHisGln 971  
 QY 1371 TCCCTGTGGCAGGATGGACGAGAGCATGAGCGGCCCG-----GGCCCCA 1324  
 Db 972 GlyAlaValGlySerProGlyProAlaGlyProArgGlyProValGlyProSerGlyPro 991  
 QY 1323 CACGGGAA-----GCAGCTGGCAGCCCTCAGGTGTCTGTCTCCACCTTGGGCTGGT 1270  
 Db 992 ProGlyLysAspGlyAlaSerGlyHisProGlyProIle--GlyProPro----- 1007  
 QY 1269 GGGGGTCAAGGTGGGACACCGATGGAGGCGAGGGGATGGAGGGGAA 1217  
 Db 1008 --GlyProArg-----GlyAsnArgGlyGluArgGlySerGlu 1019  
 RESULT 15  
 DRPL RAT STANDARD; PRT; 1183 AA.  
 ID DRPL RAT  
 AC P54258;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Atrophin-1 (Dentatorubral-pallidolysian atrophy protein).  
 GN DRPLA  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE=Cerebellum, and Striatum;  
 RX MEDLINE=97317138; PubMed=9173996;

RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,  
RA Ashworth R.G., Ross C.A.:  
RT "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)  
RT homologue";  
RL Neurobiol. Dis. 2:129-138 (1995).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;  
RX MEDLINE=96081227; PubMed=8541849;  
RA Schmitt I., Epplen J.T., Riess O.;  
RT "Predominant neuronal expression of the gene responsible for  
RT dentatorubral-pallidoluysian atrophy (DRPLA) in rat";  
RL Hum. Mol. Genet. 4:1619-1624 (1995).  
CC -1- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH  
CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.  
CC -1- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES  
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U31777; AAA80337.1; --  
CC EMBL; X89453; CAA61623.1; --  
CC InterPro; IPR002951; Atrophin.  
CC Pfam; PF03154; Atrophin-1; 2.  
CC PRINTS; PR01222; ATROPHIN.  
CC DOMAIN 165 171 POLY-PRO.  
CC DOMAIN 303 306 POLY-PRO.  
CC DOMAIN 377 383 POLY-SER.  
CC DOMAIN 387 391 POLY-SER.  
CC DOMAIN 440 446 POLY-PRO.  
CC DOMAIN 477 480 POLY-HIS.  
CC DOMAIN 481 489 POLY-GLN.  
CC DOMAIN 502 505 POLY-PRO.  
CC DOMAIN 562 572 POLY-SER.  
CC DOMAIN 702 705 POLY-PRO.  
CC CONFLICT 455 455 N -> S (IN REF. 2).  
CC CONFLICT 594 594 F -> L (IN REF. 2).  
CC CONFLICT 689 689 P -> R (IN REF. 2).  
CC CONFLICT 717 717 T -> M (IN REF. 2).  
CC CONFLICT 737 737 A -> V (IN REF. 2).  
CC CONFLICT 965 965 MISSING (IN REF. 2).  
CC SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9BIF CRC64;  
  

Alignment Scores:  
Pred. No.: 6e-07 Length: 1183  
Score: 304.50 Matches: 327  
Percent Similarity: 28.73% Conservative: 95  
Best Local Similarity: 22.26% Mismatches: 437  
Query Match: 3.23% Indels: 610  
DB: 1 Gaps: 71

US-09-931-704-3 (1-5087) x DRPLA\_RAT (1-1183)

Qy 1090 CCAAGACTCTCTCAGCTGACACACAG-----ATCATCTCAAGTATCT 1134  
Db 120 ProArgAspIleAspGlnAspAsnArgSerThrSerProSerIleTyxSerProGlySer 139  
Qy 1135 ACTGATAGACATCTCGCTGCGCAAGTCCTCATCTCAACATACACATGCTCTCTTC 1194  
Db 140 ValGluAsnAspSer-----AspSerSerGlyLeuSerGln 152  
Qy 1195 TCTCCCGTCTTCCAGGAGTGTTCCTCCCTCCATCCCTCTGCTCCATCTGCTGTC 1254  
Db 153 GlyProAlaArgProTyxHisProProLeuPhePro-----ProSerProPro 169  
Qy 1255 CCACCC-----TCACCCCAACCCAGCCAGGTGGGGACAGACACCTGAGGGCTGCCAGC 1311

Db 170 ProProAspSerIleProArgGlnProGluSerGlyPheGluPro----- 184  
Qy 1312 TGCCTTCCCGTGTGGCGCGCGCTCATGTCTTCGTCCATCTCTGCCCACAGGGA 1371  
Db 185 -----HisPro----- 186  
Qy 1372 CTGCTGGGGATGTTAGCTGCTCT-----GTGACGGTGTCTTGGCACTCTCCCTGC 1422  
Db 187 -----SerValProProThrGlyTyxHisAlaProMetGluProProThr 201  
Qy 1423 AGTCCAGCTCTCAATCGCACAGGGGCCAGGGCTGGCCCTCCATCCAGAAACCTA 1482  
Db 202 SerArgLeuPheGlnGly----- 207  
Qy 1483 TGACCTCACCGCTACCTGGAGCACCACTCCGAGCTTGGCTGGGACCTATGTAGTAT 1542  
Db 208 -----ProProProGlyAlaProPro-----His 216  
Qy 1543 CCAGCTAGGAATCTGGAGTTGGGAGGAGTAGGAGTTGGGAAAGACAGTCTTACC 1602  
Db 217 ProGlnLeuTyxProGlySerAlaGlyGly-----GlyValLeu----- 229  
Qy 1603 GTGGAGGTTCTGCTAATGATGGGTGAGGAGGG----- 1638  
Db 230 -----SerGlyProProMetGlyProLysGlyGlyAlaAlaSerSerValGly 246  
Qy 1639 ---CTCTTTGGCTCCCACTCCCTGTCTGTGTCTATCTCTGCTCTCTCTCTCTTAGG 1695  
Db 247 ProProSerGlyGlyLysGlnHisPro----- 255  
Qy 1696 TGGCCCCCCTCTCCCATCTCCCTGGCCCCAGGACTAGGATGGGAGGCTCGCACC 1755  
Db 256 ---ProProThrThrProIleSerSerGlyAlaSerGlyAlaPro----- 272  
Qy 1756 CGCCTTGGCCCATTTGCCCACTGGCTGCCAGCCAGCCCGCTCCCTCCCTGGGGCGC 1815  
Db 273 -----ProAlaLysProProAsnThrProValGlyAla 283  
Qy 1816 GGGAGTCTCCT-----CTGTTTACACCTGTTGTGTGTCTCTCTTC 1857  
Db 284 GlyAsnLeuProSerAlaProProAlaThrPhePro-HisValThrProAsnLeu-- 302  
Qy 1858 CGGGCGGGGTGGTGGGGACAGAGGGCCCTCCATGCTCTGCTGCTCCAG----- 1912  
Db 303 -----ProProProAlaLeuArgProLeu 311  
Qy 1913 ---CTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1968  
Db 311 uAsnAsnAlaSerAlaSerProProGly-----MetGlyAlaGlnProIleProGly 328  
Qy 1969 CGCTCTGCTCTCCCATCTCTAGTGGCTCTTAGGGGGTCTAGGGGAGAGGGAGTGT 2028  
Db 328 yHisLeuProSerProHis-----AlaMetGlyGlnGlyMetSe 341  
Qy 2029 AGGAAACCCAGGACAGTGTGGAGGGGTTAGGTGTGGATGGAGGTATCTGTAGG 2088  
Db 341 rGlyLeuPro----- 344  
Qy 2089 ATTTGGGGTGTCTCCAGAGTGTTCAGAGAGCCAGGAGAGAGAGAGGGTGGAGG 2148  
Db 345 -----Pro-GlyProGluLysGlyProThrLeu 354  
Qy 2149 AGCCGAGGACCATCGGGAACCGGCCCTCTTCCGCTGCTCTTCCATATCCAGAG- 2207  
Db 354 laProSerProHisProLeuProProAlaSerSerAlaProGlyProProMetArgT 374  
Qy 2208 ---CCCTACTCTCGACCCAGGNA-----AGAAAGGAGAGAGGTGGCGG 2250  
Db 374 yrProTyxSerSerCysSerSerSerValAlaAlaSerSerSerSerAlaAla 394  
Qy 2251 GGGAGCTGGCTCCAGCC-----CCAGGATACACCGAGGAAATAGTTTCTCTC 2298  
Db 394 hrSerGlnTyxProAlaSerGlnThrLeuProSerTyxProHisSerPhe----- 410

QY 2299 TGTCTTTGTGACGCTGTGAACCTCCCTCCCTGGCCCTTGCCTATCCAGGCTCTCCCTT 2358  
Db 411 -----ProProProThrSerMetSerValSerAsnGlnProProL 424  
QY 2359 GCTTC-----TCCCTCTTTCCAGTATATACATCCCTCATCCCTTTCCCTGGC 2409  
Db 424 ystYrThrGlnProSerSerGlnAlaValTpsSer-----GlnGlyp 440  
QY 2410 CCCAGCCGCTCCCGGAGGTTGAAAGGCTCTGCCCTCTCCCTATACCATGCTGCT 2469  
Db 440 roProProProProProTyroGlyArgLeuLeuProAsnAsnThrHisPro----- 458  
QY 2470 TCCATAGCTTCCCTCTCTACTCATGAGACTGCCTCCATCTTCTCTCTCT 2522  
Db 459 -----GlyProPheProProThrGlyGlyG 467  
QY 2523 --GCAACCTTCTCTATCAGTGAACCTCTTTCGGAGTGTAGTACGACCGTCTC 2580  
Db 467 lnSerThrAlaHisProProAlaProAlaHisHisHisGlnGlnGlnProG 487  
QY 2581 TCCCCAGCCCTCAGCTGGTGGGCTGGTGTGTGTCAGCGGCAATGGGCTCTGTTCCA 2640  
Db 487 lnProGlnProGln-----ProGlnGlnHisHisHisGlyAsn 500  
QY 2641 ATGGGCACTCTCATCTCTCTCTTCTTCTTCTGTCAGAAAACCTTGTCTCACTCCA 2700  
Db 500 erGlyPro-----ProProProGlyAlaValProHisP 511  
QY 2701 CCTCTCTAGTCCCGACCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760  
Db 511 roLeuGluSerSerAsnSerHisHis-----HisProTyroAsnMetSerProSerL 529  
QY 2761 GTGGT---CTACACCT 2817  
Db 529 euGlySerLeuArgProTyroProProGlyProAlaHis----- 541  
QY 2818 CAATCTGGCTTCAGGCCCCAGCAATGGTCTCTCTCAAGTCTGTCAGCACTCTCTG 2877  
Db 542 -----LeuPro-----ProSerHisGlyGlnValSerTyroSerGlnAlaGlyPro----- 556  
QY 2878 AAGCCCGACAGTGTTTTGAAGGCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2937  
Db 556 ----- 556  
QY 2938 GAGCGTGTGCTTCTCGAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2996  
Db 557 -----AsnGlyProProV 561  
QY 2997 --TCTACTCTCAGCTCTCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3048  
Db 561 alSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerHis- 580  
QY 3049 GCGGCACTCTCCCAAGTTTGGCCAGCCCAATCAGCAGCTCTCTCTCTCTCTCTCTCT 3108  
Db 581 -----ProSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 592  
QY 3109 GTGCGTCT 3168  
Db 592 yrProPheProProValPro-----Pro-IleThrThrSerSer 604  
QY 3169 GCTTCACTGTCACTTCATCAATATATATCTTATTTGAAATACTCAGGAGGCCATG 3228  
Db 605 AlaThrLeuSerThrValIleAlaThrValAla----- 615  
QY 3229 AACAAAGAGCTTAGCATGGAGAGCGGCGAGTGTTCAGGGGACACAAAAATAGAACTT 3288  
Db 616 -----SerSerProAlaGlyTyroLys----- 622  
QY 3289 TGGGACAGGTATCTCTCTGGTGTGAGCCAGCGGCTCTGCCCTCTCTCTCTCTCTCT 3348  
Db 622 ----- 622

QY 3349 CCTCTCTTTTTCACAGCTGAACCTACCTGGGCCCCCTTTTCAACGAGCAGACTTCAACCC 3408  
Db 623 -----ThrAlaSerProProGlyProPro-----GlnTyroSerLysArgAlaPro 637  
QY 3409 TCC-----CCGCTGGGGGAGAGACTCTCTGCCAGGCGCAC 3444  
Db 638 SerProGlySerTyroLysThrAlaThrProProGlyTyroLysProGly----- 653  
QY 3445 TGTTCAGCTGGAGGTGGCGAAGCTCAATGACAACTCGCGCTGACCCAGAACTACGA 3504  
Db 654 -----SerProSerPheArgThrGlyThrProProGlyTyroArg 667  
QY 3505 GGCCTACAGCCACTTCTGTGTCTGCTGGCTCAACCGTCAGGCTGCCACTGCTGA 3564  
Db 668 GlyThrSerProPro-----AlaGlyPro-----GlyThrPheLysPro 680  
QY 3565 GCTGGCGCGCAG-----CCTGGCCACTTTCACAGCTCCAGGCTCGTCTGGGCG 3618  
Db 681 GlySerProThrValGlyProGlyProLeu-----ProProAlaGlyProSer----- 696  
QY 3619 CATTCGGGCGCTCATGGCAGCTCTGGGTACCCACTGCCCGCCAGCCGCTGCTGGGACTGA 3678  
Db 697 -----SerLeuSerSerLeuProProProProAlaAla----- 707  
QY 3679 ACCACTTGGACTCTGGCCCTGCCACAGTGAATCTCTCCAGAGATGGAGCTTCTG 3738  
Db 708 -----ProThrThrGlyProProLeu 714  
QY 3739 GCTGTGAAGAGCTGCAGACCTGGCTGGCGCTCGGCCAAGACTTCAACCGGCTCAA 3798  
Db 715 ThrAlaThrGlnIleLysGln-----GluProAlaGlu 725  
QY 3799 GAAGAAGATGCCAGCTCCAGCAGCTGCAGTACCCTGCACCTGGGGGCTCATGGCTCTG 3858  
Db 726 GluTyroGluThrProGluSer-----ProValProPro----- 736  
QY 3859 ACTTCTGACT 3918  
Db 737 -----AlaArgSerProProProProProLysValValAspVal 749  
QY 3919 GCCCTGTATGCAACACCTGTGTGAGCCAGGACAGAGCTGTGAGCTCTGCCCCCTTC 3978  
Db 750 ProSerHisAlaSerGlnSer-----AlaArgPheAsnLys-----His 762  
QY 3979 CTGACCCGCTGGCGTGTGATGCGATCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4038  
Db 763 LeuAspArgGlyPheAsnSerCysAlaArgSerAspLeuTyroPheValProLeuGluGly 782  
QY 4039 ACCAGCTGGGAGAGGTACAGTAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4098  
Db 783 SerLysLeuAlaLysLysArgAlaAsp-----LeuValGlu-LysVa 796  
QY 4099 CTCGAGGAGGTGAA----- 4114  
Db 796 lArgArgGluAlaGluGlnArgAlaArgGluLysGluArgGluArgGluArg 816  
QY 4115 -----GTGTTTCAGGTTCTGGTGC 4131  
Db 816 gGluLysGluArgGluArgGluLysGluArgGluLysValLysLeuAlaG 836  
QY 4132 AGAGCGCTCATGGCT 4191  
Db 836 nGluGly----- 838  
QY 4192 CAGGTGGCACATCTGGAGGCGAGGTTGAG-----GGGCCACCCACACAT 4239  
Db 839 -----ArgAlaProValGluCysProSerLeuGlyProValProHisAr 853  
QY 4240 GCCTTCTTGGGGTGAAGCCCTTTGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4297  
Db 853 gPro-----ProPheGluProGlySerAlaValAlaThrVa 865  
QY 4297 ----- 4297

Db 865 lProTyrLeuGlyProAspThrProAlaLeuArgThrLeuSerGluTyrAlaArgPr 885  
Qy 4298 -----TTATCCCAATCACTTATACATCAATTCAGGAAACAAACATGGTGGCAAT 4350  
          :::|||||:::|||||  
Db 885 ohisValMetSerProGlyAsnArgAsnHisPro-Phe----- 897  
Qy 4351 TCTACAAAAAGAGATGAGATTAAACAGTGCAGGGTTGGGTCTGCAATTGGAGTGCCT 4410  
          |||:::|||||  
Db 898 --TyrValProLeuGlyAlaValAspProGlyLeuLeuGlyTyrAsnValProAlaLeuT 917  
Qy 4411 ATAAACACAGAGAGAAAATACTGAAAGCAGAGGGGAGGACAGACACAGACAGACCCAG 4470  
          ||:::|||||  
Db 917 yrSerSerAspProAlaAlaArgGluArgGluArgGluAlaArgGluArgAspLeuArgA 937  
Qy 4471 GAGTCTCAAGACACAGAGTGGCAAAACCCGAGCTGAGCATCAGGACCTTGCTCG 4530  
          |||  
Db 937 spArgLeuLys-----ProGlyPheG 944  
Qy 4531 AATTGCTTCCAGTATTACGGTGCCTCTTCTCTGCCCTTTCCAGGGTATCTGTGGT 4590  
          ||:::|||||  
Db 944 luValLysProSerGluLeuGluProLeu-----HisGlyV 956  
Qy 4591 TGCCAGGCTGGGAGGGCAACCATAGCCACACACAGGA----- 4629  
          :::|||||  
Db 956 alProGlyProGlyLeuAspProPheProArgHis-GlyGlyLeuAlaLeuGlnProGly 975  
Qy 4630 -----TTCTCTGAAAGTTTACATGCAGTAGCATTTTGGGGTGTAGG 4671  
          |||  
Db 976 ProProGlyLeuHisProPheHisProSerLeuGlyProLeuGluArgGluArg 995  
Qy 4672 GTGGCAGCTCCCAAGGCGCTGCCCGCCAGCCGCCACTCATGACTCTAAGTGTGTG 4731  
          :::|||||  
Db 996 LeuAlaLeuAlaGlyProAlaLeuArgPro----- 1006  
Qy 4732 TATTAATATTATTATTGGAGATGTTATTATTATAGATATTTATTGC---AGAATT 4788  
          |||  
Db 1007 -----AspMetSerTyrAlaGluArgLeu 1014  
Qy 4789 TCT-----ATCTTGTTATTAACAATAAATGCTTGCCCA 4824  
          :::|||||  
Db 1015 AlaAlaGluArgGlnHisAlaGluArgValAlaAlaLeuGlyAsnAspProLeuAlaArg 1034  
Qy 4825 GAACTTAGTCTCTTTGCCAGCTCACCCCTCCTGGTGTCTCATCAGACTCTTGCCACCCC 4884  
          |||  
Db 1035 LeuGlnMetLeuAsnValThrProHisHisHisGlnHisSerHisleHisSerHisLeu 1054  
Qy 4885 TGGCTCCCACTCCCTGCTTGCCTCTCTGGTGGAGCTGCACAGAGCTCTGGGAAGAGGCCCTC 4944  
          |||  
Db 1055 HisLeu-HisGlnGlnAspAlaIleHisAlaAla-----SerAlaSerValHisProLe 1072  
Qy 4945 TTCCTCCCGCACTGGGGCGATGGGCGCACCTCAGACTTACCCACTGCTGCTGCCACC-- 5002  
          |||  
Db 1072 ulleAspProLeuAlaSerGlySerHisleuThrArgIleProTyrProAlaGlyThrLe 1092  
Qy 5003 -ACCAACCCCTTGATCCCTCAGTCCCTCCACACAGCTTCTGTG-----CACCCAGGTT 5055  
          |||  
Db 1092 uProAsnProLeuLeuProHisProLeuHisGluAsnGluValLeuArgHisGlnLeuPh 1112  
Qy 5056 TCCCTCACCCAC 5068  
          ||  
Db 1112 eAlaAlaProTyr 1116

Search completed: January 27, 2003, 16:05:35  
Job time : 244.026 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:48:57 ; Search time 318.364 Seconds  
(without alignments)  
6584.667 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 9432

Sequence: 1 aactcgagtgaggcctggc.....ccttgcgaagtctctcca 5087

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09931704/runat\_27012003\_154126\_3635/app\_query.fasta\_1.7189  
-DB=SPTREMBL\_21 -QMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPWT=spt -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09931704@cgn\_1\_1\_633 @runat\_27012003\_154126\_3635 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	896	9.5	225	4 Q9UBD9	Q9ubd9 homo sapien

2	863	9.1	225	11	Q9QZM3
3	420	4.5	5146	6	Q8SPM4
4	404	4.3	4123	4	O75851
5	397.5	4.2	1315	10	Q9SPM0
6	396.5	4.2	763	2	Q9XDH2
7	382.5	4.1	1047	5	Q9VR13
8	378.5	4.0	2284	5	Q9VPG1
9	374.5	4.0	1480	10	Q9LIE8
10	373	4.0	676	6	Q95JC9
11	370.5	3.9	1548	4	Q9NYW9
12	370.5	3.9	2161	4	Q9Y566
13	369	4.0	1458	13	Q910B9
14	358.5	3.8	1616	4	O15054
15	354	3.8	2158	11	Q9WU13
16	351.5	3.7	2087	11	Q9WU88
17	351	3.7	2167	11	Q9WV48
18	341.5	3.6	2703	5	Q9VEG7
19	339.5	3.6	1491	13	Q91718
20	336.5	3.6	1003	4	Q96DN6
21	336	3.6	1188	10	Q41805
22	336	3.6	2715	5	O61603
23	333.5	3.5	967	4	Q96Q00
24	333	3.6	774	12	O41971
25	333	3.6	1414	5	Q26634
26	333	3.5	3084	12	Q8UZ11
27	331.5	3.6	1486	13	Q91717
28	329.5	3.5	1289	10	Q9FLQ7
29	329.5	3.5	3570	4	Q99552
30	327.5	3.5	706	12	O41972
31	327	3.5	1418	13	Q9W7R9
32	326.5	3.5	2157	4	Q95875
33	326	3.5	1953	5	Q9BIT7
34	325.5	3.5	1182	4	Q99495
35	325	3.5	1449	13	Q910C0
36	323	3.4	2187	11	P70670
37	321.5	3.4	1608	4	Q96RKO
38	320	3.4	2157	4	Q96QC6
39	319.5	3.4	1006	10	Q9LMQ1
40	319	3.4	1453	4	Q9Y6T1
41	318.5	3.4	2971	4	Q9Y5L9
42	318	3.4	1378	5	Q97405
43	315	3.3	566	6	Q95JDI
44	314.5	3.4	1447	13	Q9IB91
45	314	3.4	1487	4	Q14047

## ALIGNMENTS

RESULT 1

ID	Q9UBD9	PRELIMINARY;	PRT;	225 AA.
AC	Q9UBD9;			
DT	01-MAY-2000	(TREMBLrel. 13, Created)		
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	Neurotrophin-1/B-cell stimulating factor-3 (Cardiotrophin-like cytokine)	(Similar to cardiotrophin-like cytokine, neurotrophin-1/B-cell stimulating factor-3).		
DE	neurotrophin-1/B-cell stimulating factor-3.			
GN	CLC.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=9432254; PubMed=10500198;			
RA	Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S.,			
RA	Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,			
RA	Simonet W.S., Boone T., Chang M.-S.			
RT	"Novel neurotrophin-1/B cell-stimulating factor-3: A cytokine of the IL-6 family."			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).			
RN	[2]			

```
RP SEQUENCE FROM N.A.
RX MEDLINE=9382254; PubMed=10448081;
RA Shi Y., Wang W., Yourey P.A., Gohari S., Zukauskas D., Zhang J.,
RA Ruben S., Alderson R.F.;
RT "Computational EST database analysis identifies a novel member of the
RT neuroplectic cytokine family.";
RL Biochem. Biophys. Res. Commun. 262:132-138(1999).
RN (3)
RP SEQUENCE FROM N.A.
RP Hu X., Xu Y., Zhang B., Peng X., Yuan J., Qiang B.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL (4)
RP SEQUENCE FROM N.A.
RP TISSUE=KIDNEY;
RA Strausberg R.;
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176912; AAF00992.1; -
DR EMBL; AF172854; AAD54284.1; -
DR EMBL; AF176911; AAF00991.1; -
DR EMBL; AY049775; AALJ5436.1; -
DR EMBL; BC012933; AAH12939.1; -
SQ SEQUENCE 225 AA; 25176 MW; E2DD4B6280833B55 CRC64;

Alignment Scores:
Pred. No.: 1-96e-55 Length: 225
Score: 896.00 Matches: 170
Percent Similarity: 90.72% Conservative: 6
Best Local Similarity: 87.63% Mismatches: 12
Query Match: 9.50% Indels: 2
DB: 4 Gaps: 2

US-09-931-704-3 (1-5087) x Q9UBD9 (1-225)
QY 3311 GGTGAGCGAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCTTTTACAGCTGAACCTACCTGGCGCCCTTTCAACGAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCAACCTCCCGCTGGGGCAGAGACTTGCACGAGGCGACTGTGACTTG 3454
Db 72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91
QY 3455 GAGGTGGCGAGCCTCAATGACAACTCGGCTGACCCAGAACTACGAGGCGCTACAGC 3514
Db 92 GluValTyrArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGTGGCTCAACCGTCAAGCTGCGCTGCTGAGCTGCGCGC 3574
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeuArgArg 131
QY 3575 AGCTGCGCCCACTTCTGCACAGCCTCCAGGCTGCTGGCAGCATTTGGCGGCTCATG 3634
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGTACCACTGCGCCCGCTGCTGGAGCTGAACCACTTGGACTCCT 3694
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTyrPro 171
QY 3695 GGCCTGCCACAGTACTTCTCCAGAAGATGGACGACTTCTGGCTCTGAAGAGAGCTG 3754
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTyrLeuLeuLysGluLeu 191
QY 3755 CAGACTGGCTGGCGCTGGCCAGGACTTCAACGGGCTCAGAGAGATGAGGCT 3814
Db 192 GlnThrTrpLeuTyrArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211
QY 3815 CCAGCAGCTGAGCACCCTGACCTGGGGCTCATGGCTTC 3856
Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 2
Q9QZM3 PRELIMINARY; PRT; 225 AA.
AC Q9QZM3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Neurotrophin-1/B-cell stimulating factor-3.
GN BSF3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432254; PubMed=10500198;
RA Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S.,
RA Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,
RA Simonet W.S., Boone T., Chang M.-S.;
RT "Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the
RT IL-6 family.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).
DR EMBL; AF176913; AAF00993.1; -
DR MGD; MGI:1930086; Bsf3.
SQ SEQUENCE 225 AA; 25261 MW; 68B1FEAAB7F1A950 CRC64;

Alignment Scores:
Pred. No.: 4-38e-53 Length: 225
Score: 863.00 Matches: 163
Percent Similarity: 88.66% Conservative: 9
Best Local Similarity: 84.02% Mismatches: 10
Query Match: 9.15% Indels: 12
DB: 11 Gaps: 2

US-09-931-704-3 (1-5087) x Q9QZM3 (1-225)
QY 3311 GGTGAGCGAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCTTTTACAGCTGAACCTACCTGGCGCCCTTTCAACGAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCAACCTCCCGCTGGGGCAGAGACTTGCACGAGGCGCTGTGACTTG 3454
Db 72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAsnLeu 91
QY 3455 GAGGTGGCGAGCCTCAATGACAACTCGGCTGACCCAGAACTACGAGGCGCTACAGC 3514
Db 92 GluValTyrArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGTGGCTCAACCGTCAAGCTGCGCTGCTGAGCTGCGCGC 3574
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeuArgArg 131
QY 3575 AGCTGCGCCCACTTCTGCACAGCCTCCAGGCTGCTGGCAGCATTTGGCGGCTCATG 3634
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGTACCACTGCGCCCGCTGCTGGAGCTGAACCACTTGGACTCCT 3694
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTyrPro 171
QY 3695 GGCCTGCCACAGTACTTCTCCAGAAGATGGACGACTTCTGGCTCTGAAGAGAGCTG 3754
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTyrLeuLeuLysGluLeu 191
QY 3755 CAGACTGGCTGGCGCTGGCCAGGACTTCAACGGGCTCAGAGAGATGAGGCT 3814
Db 192 GlnThrTrpLeuTyrArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211
QY 3815 CCAGCAGCTGAGCACCCTGACCTGGGGCTCATGGCTTC 3856
Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225
```



Db 3417 sProAlaProGlnHisGlyGlyAlaProCysProGlyGluAlaGlyGlyAlaGly 3437  
Qy 1292 ACACCTGAGG---GGCTGCGCAGCTGCTTCCCGGTG--- 1325  
Db 3437 nHisGlnArgGluThrCysAlaSerThrProGluCysProValAspGlyAlaTrpSerPr 3457  
Qy 1326 ---GGCCCG--- 1355  
Db 3457 oTrpGlyProTrpSerProCysGluValCysLeuGlyArgSerHisArgSerArgGluCys 3477  
Qy 1356 -TCCTCCACACAGGAGACTCGTGGGATGTTAGCGTCCCTGT--- 1397  
Db 3477 sSerTrpProProThrSerGluGlyGly---ArgProCysProGlyGlyHisArgGly 3495  
Qy 1398 ---GCACGGTGTCTCGGACCTCCCTCGCAGTGCAGCTCTCAATCGACAGGGGACCCA 1453  
Db 3495 nSerArgProCysGlnGlyAsnSerThrGlnCys---ThrAspCysAlaGlyGlyGly 3513  
Qy 1454 GGGCTGGCCCT--- 1466  
Db 3513 nAspLeuLeuProCysGlyGlnProCysProArgSerCysGluAspLeuSerProGlyVa 3533  
Qy 1467 ---CCATCAGAAAACCTATGACTCACCGCTACCTGGAGACCACTCCGC 1516  
Db 3533 lGluCysGlnProAspSerMetGlyCysGlnGlnProArgCys---GlyCysProProGlyG 3553  
Qy 1517 AGCTTGGCTGGGAC---CTATGTGAGTAT--- 1542  
Db 3553 lLeuSerGlnAspGlyLeuCysValThrProSerGlnCysArgCysGlnTrpGlnProG 3573  
Qy 1543 ---CCAGCGTAGGAATCTGGAGTTGGGA---GGAGTGAGGAGTTGGG 1585  
Db 3573 lYalaMetGlyIleProGluAsnGlnSerArgSerAlaGlySerGlyLeuSerSerTrp- 3592  
Qy 1586 GAAAGACAGTCTAACCGTGGAGGTCTCGTAAATCATGGGTGAGGAGGGGCTCTTTG 1645  
Db 3593 Glu---SerLeu 3595  
Qy 1646 GCTTCCC---ACCAGTCCCGCTGCTGTGTATCTCTCCCTTCCCTCTTAGGT 1696  
Db 3596 GluProGlyGluValThrGlyProCysAspAsnCysThrCysValAlaGlyIleLeu 3615  
Qy 1697 GGGCCCCCATTCCCATTCCTGGCCCCAGGACTAGGCATGTGGAGCGCTCGCACCC 1756  
Db 3616 GlnCysGlnGluValProAlaCysSer---GlyLeuGlyLeuTrpGlySerTrpGlyPro 3634  
Qy 1757 GCCTGGCCATTGCCCTGCTGCTG---CAGCCC 1789  
Db 3635 ---TrpGluAspCysSerValSerCysGlyGlyGlyGlyGluGlnLeuArgPheArgArgCys 3653  
Qy 1790 AGCGCCCGCTCCCTCGGGCGCGGGGAGTCTCCT--- 1827  
Db 3654 ProArgProProCysPro---GlyProAlaArgGlnSerArgThrCysArgThrGlnValC 3673  
Qy 1828 ---CTGTTTACCGTGT---T 1843  
Db 3673 yeArgGluAlaGlyCysProAlaGlyArgLeuTrpArgGluCysGlnProSerGluGlyC 3693  
Qy 1844 GTGGTCTCTTCCGGGGGGGTGGTGGGGACAGAGGGCCCCACCTCCCATCCCT 1903  
Db 3693 yeProPheSerCysAlaHisValThr---GlyGlnValGlyCysPheSerAlaGlyC 3711  
Qy 1904 GCGTTCAGCTCGCTCTGCCCCAG--- 1929  
Db 3711 yeGluGluGlyCysHisCysProGluGlyThrPheLeuHisArgSerAlaCysValGlnG 3731  
Qy 1930 ---ACCTGG---GGCCCTGCTGCTCTGAGCC 1954  
Db 3731 luCysProCysValLeuThrAlaLeuTrpLeuGlnGlyLeuGlyAlaAlaGlyAlaAsp 3751  
Qy 1955 CAGGGCCCTCTCCGTCTCGCTCTCCCATCTAGCTGGGCTCTCTAGGGGGTCTATGG 2014  
Db 3751 roGlyAlaHisLeu---SerValLeuG 3759

Qy 2015 GGGAA---GGGACTGTAGGGAACCCAGCAGTACTAGTGGCAGGGGTTTAGGGTGTGATG 2071  
Db 3759 lyGluAsnGlyGlnProLeu---GlyProGlyAspGlu--- 3770  
Qy 2072 GAGGTTATGCTTAAGGATTTGGGGTGTGTCCAGAGGTGTTTCAGAGAGCCAGAGAGAA 2131  
Db 3771 ---LeuGlySerGlyGlnSerLeu---ArgThr 3779  
Qy 2132 GGA--- 2137  
Db 3780 GlyCysHisAsnCysSerCysAlaHisGlyLysLeuSerCysSerValGluAlaCysSer 3799  
Qy 2138 AGGTTGGAGGAGCGAGGACCATGGGAAACCCGCCCTCTTCCCGTGTCTCTCTTCC 2197  
Db 3800 LysAlaAlaGlyGlyPheSerProTrp---GlyProTrpGlyProCysSerArgSer 3817  
Qy 2198 ACATCCACAGCCTACTCTGGAGCCAGGGAAGAAAGGAGAGGTGGCGGGGAGCT 2257  
Db 3818 Cys---GlyGlyLeu 3821  
Qy 2258 GGCTCCAGCCCCAGGATACACGAGGAAATAGTTGTCTGTGCTTGTACCGTGTGA 2317  
Db 3822 GlyThrArgThrArgSerArgGln--- 3829  
Qy 2318 ACCTCCCGCTGGGCCCTTGCCTATCCAGGCTCTCCCTTCTCTCTCTCTCTTCCCA 2377  
Db 3830 ---CysValArgProMetPro---AlaProGlyGlyGln 3840  
Qy 2378 GTTATACATCTCCCTCAT---CCCTTTCCTGGGCCCCAGCGCT 2419  
Db 3841 GlyCysHisGlyProHisTrpAspLeuGluTrpCysProSerProGluCysProGlyAla 3860  
Qy 2420 CCCCAGGGTGGAAAGGGCTCTCCCTCTCTCCCT--- 2455  
Db 3861 AlaGlySerThrAlaGluProAlaThrGlyLeuProGlyGlyTrpGlyLeuTrpSerPro 3880  
Qy 2456 ---ATACCATGC--- 2464  
Db 3881 TrpSerProCysSerGlyThrCysThrAspProAlaHisProAlaTrpArgSerArgSer 3900  
Qy 2465 ---TGCTTTCATAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2518  
Db 3901 ArgLeuCysLeu---AlaAsnCys---ThrGlyGlyAlaAlaSerGlnGluArg 3916  
Qy 2519 TTCTGCAACCT 2578  
Db 3917 ProCysAsn---LeuProSerCys---ThrGluLeuProLe 3928  
Qy 2579 TCTCCCGAGCCCTCAGCTGTGGCTGGGTGTCTCAGCGGC---AAATGGGGCTC 2632  
Db 3928 uCysPro---GlyProGlyCysGluAlaGlyAsnCysSerTrpThrAl 3943  
Qy 2633 TGGTTCCAAATGGGCACCT 2692  
Db 3943 aTrpAlaProTrpGluProCysSerArgSerCysGlyValGlyGlnGlnArgLeu--- 3962  
Qy 2693 CTCACCTCCCTCTCTAGTCTCCGACCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2752  
Db 3963 ---ArgAlaTrpHisProProGly--- 3969  
Qy 2753 TCCAAGGAGTGGTCTACACCTCTGCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 2812  
Db 3970 ---Pr 3970  
Qy 2813 CCCTGCAATCTGGCTTCCAGGCCCCAGCAATGGTCTCTCAGAGTCTGTCAGGACCTCC 2872  
Db 3970 oGlyGlyHisTrpCysProAspValLeuThrAlaTrpGlnGluArgArgPhe--- 3987  
Qy 2873 TTGCCAAGCCCGACAGTGTGTTGAAGGCTCATCTCTCTCTCTCTCTCTCTCTCTCTCT 2932  
Db 3988 -CysAsnLeuArgAlaCys--- 3993



Db 4630 heArgLeuArgTrpArgGluAlaGlyLeuProProGlyGlyGlyCysArgGlyProTrpA 4650  
QY 4405 -----TGCCCTATTAAACAGAGAGAAATACTGAA 4435  
Db 4650 laGlnThrGluSerCysAsnMetGlyProCysPro----- 4661  
QY 4436 AGCACAGGGCAGGGAC-----AGACCAGACAGACCCAGG 4471  
Db 4662 -----GlyGluSerCysGluAlaGlnAspThrValProThrProAspCysAlaAsnG 4679  
QY 4472 AGTCTCCAAAGCACAGAGTGCACAAACACCCAGCTGAG----- 4512  
Db 4679 inCysProArgSerCysValAspLeuTrpAspArgValGluCysLeuGlnGlyProCysA 4699  
QY 4513 -----CATCAGGACCTTGCCCTCGAATTGCTTC 4540  
Db 4699 rgProGlyCysArgCysProProGlyGlnLeuValGlnAsp-----GlyHisCysValP 4717  
QY 4541 CAGTATTACGGTGCTCTTCTGCCCCCTTCCAGGGTATCTGTGGTTGCCAGGCTG 4600  
Db 4717 roValSerSerCysArgCysGlyLeuProSerProAsnAlaSerTrpAlaLeuAlaProA 4737  
QY 4601 GGCAG-----GGCAACCATAGCCACACAGGATTTCTGAAAGTT 4642  
Db 4737 laGluValValArgLeuAspCysArgAsnCysThrCysValAsnGly-----SerL 4754  
QY 4643 TACAATGCAGTAGC-----ATTTTGGG----- 4665  
Db 4754 euAlaCysSerSerHisGluCysProThrLeuGlyProTrpSerAlaTrpSerAsnCysS 4774  
QY 4666 -----TGTAG-GGTGGCAGCTCCCAAGG----- 4698  
Db 4774 exAlaProCysGlyGlyThrThrLysArgHisArgSerCysLysGluGlyProGlyV 4794  
QY 4689 -----GlyGlnProCysGlnAlaGlnAspMetGluGlnGlnAspCysAsnLeuProCysP 4814  
QY 4696 CCCAGCCCCCACCACCTCATGAC-----TCTAAGTGTTGTGTTAATAATATTATTATTATTGG 4752  
Db 4814 roGluCysProProGlyGlnValLeuSerAlaCysAlaVal----- 4827  
QY 4753 AGATGTTATTATTAGATGATATTATTGCGAATTTCTATTCTGTATTAAACAAATAAA 4812  
Db 4827 ----- 4827  
QY 4813 ATGCTTGCCCGCAACTTAGTCTCTT----- 4838  
Db 4828 --SerCysProArgLeuCysSerHisLeuGlnProGlyThrProCysMetGlnGluProC 4847  
QY 4839 -----TGCCCGACCTCACCCCTCTGCTGCTCATCAGACTCTTGCCAC 4881  
Db 4847 ysGlnLeuGlyCysAspCysProArgGlyGlnLeuLeuHisAsnGlyThrCysValProp 4867  
QY 4882 CCTGTGCTCCACTCCCTGCG-----TTGCTCTGG 4911  
Db 4867 ro---AlaGluCysProCysThrGlnLeuSerLeuProTrpGlyLeuThrLeuThrLeuG 4896  
QY 4912 TGGAGCTGCACAGAGCTCTGGGAAGAGGCGCTCTTCTCTCCCGCACGTGGCGCATGGCG 4971  
Db 4886 luGluGlnHisArg-----GluLeuProPro-----GlyThrLeu-Leu 4898  
QY 4972 CACCTCAGACTTACCACTGCTGCTGCCACACCAACCCCTTGATCCCTCAGTCCCTCCCA 5031  
Db 4899 ThrGlnAsnCysThrHisCysIleCysGlnGlyAlaPheSerCysSerLeuThrAsp 4918  
QY 5032 CACAGCTTCTGTCACCC 5049  
Db 4919 CysGlnGluCysProPro 4924  
RESULT 4  
O75851  
ID O75851 PRELIMINARY; PRT; 4123 AA.

AC O75851;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE WUGSC:H DJ0751H13.1 protein (fragment).  
GN WUGSC:H DJ0751H13.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_taxid=9606;  
RN [1]  
SEQUENCE FROM N.A.  
RA Leonard S., Graves T., Strommatt C.;  
RT "The sequence of Homo sapiens PAC clone RP4-751H13";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004877; AAC36301.1; -;  
DR HSSP; P01130; 1AJJ.  
DR InterPro; IPR000923; BlueCu\_1.  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR000421; FAS8\_C.  
DR InterPro; IPR001092; HLH Basic.  
DR InterPro; IPR002223; Kunitz BPTI.  
DR InterPro; IPR002172; LDL\_recept\_A.  
DR InterPro; IPR002919; TIL\_Cysrich.  
DR InterPro; IPR000884; TSPI.  
DR InterPro; IPR001007; VWF\_C.  
DR InterPro; IPR001846; VWF\_D.  
DR Pfam; PF00754; F5\_F8\_type\_C; 1.  
DR Pfam; PF00057; ldl\_recept\_a; 11.  
DR Pfam; PF01826; TIL; 5.  
DR Pfam; PF00090; tsp\_1; 14.  
DR Pfam; PF00094; vwd; 3.  
DR PRINTS; PR00261; LDLRECEPTOR.  
DR SMART; SM00231; FAS8C; 1.  
DR SMART; SM00192; LDLA; 10.  
DR SMART; SM00209; TSPI; 14.  
DR SMART; SM00214; VWC; 1.  
DR SMART; SM00216; VWD; 3.  
DR PROSITE; PS00280; BPTI\_KUNITZ\_1; UNKNOWN\_1.  
DR PROSITE; PS00196; COPPER\_BLUE; UNKNOWN\_1.  
DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; UNKNOWN\_1.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
DR PROSITE; PS01209; LDLRA\_1; 9.  
DR PROSITE; PS00068; LDLRA\_2; 18.  
DR PROSITE; PS50092; TSPI; 22.  
KW Glycoprotein.  
FT NON TER  
SQ SEQUENCE 4123 AA; 434981 MW; 7AAB6F8DCE012FB CRC64;

Alignment Scores:  
Pred. No.: 4,11e-20 Length: 4123  
Score: 404.00 Matches: 514  
Percent Similarity: 27.68% Conservative: 142  
Best Local Similarity: 21.69% Mismatches: 711  
Query Match: 4.28% Indels: 1006  
DB: 4 Gaps: 131

US-09-931-704-3 (1-5087) x O75851 (1-4123)

QY 18 GCGGATGGGATTATTAAAGCTTCGCCGAGCGCGGCTCGCCCT----- 62  
Db 1309 GlyValTrpAspCysProAsp-----GlyAlaAspGluGlyProGlyHisCysProLeu 1326  
QY 63 -----CCCACTCCGCGAGCTCCGCGAGAGCGCGACCGCGCGCGCGAGCCCC 113  
Db 1327 ProSerLeuProThrProProAlaSer-----ThrLeuProGly---ProSerPro 1342  
QY 114 AGCCCCATGGACCTCCGCGAGAGGTGAAACCCAAACTAGCCCTGCTCTTCATACATGA 173

Db 1343 GlySerLeuAsp----- 1346  
QY 174 CAAGCAGGCCCATCTGATACCTAAACCGACGAAGTCACAGCCCTCCAACTCACTCTCT 233  
Db 1347 -----ThrAlaSerSerProLeuAlaSerAlaSer 1356  
QY 234 GCCTGCCCA-----GACCTCACCACATCC 257  
Db 1357 ProAlaProProCysGlyProPheGluPheArgCysGlySerGlyGluCysThrProArg 1376  
QY 258 TTGTGG-----ACTCAACCTCAACCGCACTAAATCAACAAATCCCAAGCTTAAACTA 311  
Db 1377 GlyTrpArgCysAspGlnGluAspCysAlaAspGlySerAspGluArgGly----- 1394  
QY 312 ATCTGAACCTTTAAAGTAACCCAGCTCTTAAACCTAACCTAGCCCAATGCAATTATAT 371  
Db 1395 -----CysGlyGlyProCysAlaProHisAlaProCysAla----- 1407  
QY 372 CTACCTAGCCAAACCTTAAGCTTTGCGCAGTCCAAAGTG-----TCCACTGAATCC 425  
Db 1408 -----ArgGlyProHisCysValSerProGluGlnLeuCysAspGlyValArgGln 1424  
QY 426 TCACCTTGCTCTCACTGAATAATCCAGAAAGCATATTTCCCACTGCCACATCCCTC 485  
Db 1425 CysProAspGlySerAspGluProAsp-----AlaCys 1436  
QY 486 CTTACAGCACCACCTCGGCTCTGGACTCTGTGTATCTCTGGATGTCCTCAACTCTGCA 545  
Db 1437 ValGluAlaProAlaPro-----ProAlaMeta\*rglyProProGly----- 1450  
QY 546 GTGCCATCAGCAACAACCGACTCGTCAATGACACTCTCTCTCTCTCTCTCTCTCTCT 605  
Db 1451 -----GlnAlaGlyProThrSerSerArgAlaPro-SerProProSerProProG 1468  
QY 606 CTTGAGGCTGATGAAGGCTCAATGAAGTCCAACTTTTCCCACTTAACCAACCAAGAA 665  
Db 1468 uAlaGlnGlyGluGlyArgGlyGlnGluArgSerArgThr-HisLeuThrValProA 1488  
QY 666 CGGGTGAACTCCACACTGCCACCGTCTCCCTGAGAGTGAGGACTAAATCTCTCTCAATC 725  
Db 1488 laGly---SerThrGlnLeuPro----- 1494  
QY 726 TAAACCCACCTTACACTTCCACACTCAGGAATCACATCTAGATAATACCAAACTAA 785  
Db 1495 -----LeuCysProGlyLeuP 1500  
QY 786 GCCCATAAGGC---AGCCCGACCTAGTGGTCTAACCTATACCTTG----- 830  
Db 1500 heProCys-GlyValAlaProGlyLeuCysLeuThrProGluGlnLeuCysAspGlylle 1519  
QY 831 -----CTTCCTATGGTGAGTCTGTTCTTGGCGCGCTCTCTCTCTCTCTCTCTCT 884  
Db 1520 ProAspCysProGlnGlyGluAspGluLeu----- 1529  
QY 885 AGAGCTGACTGTGCTCAGCTGCCAGCTCTG----- 915  
Db 1530 -----AspCysGly-GlyLeuProAlaLeuGlyGlyProAsnArgThrGlyLeuProC 1547  
QY 916 -----ACATGTGTCTTCCCACTCTGAGCTCCCTCAAGCTGCACTGGGAGCTG 965  
Db 1547 sProGluTyThrCys----- 1552  
QY 966 GAAGACTGGCAGGAGCTAGGTACAACTGGAAACACAGCAGGTGCACTGCTGCTCCCTA 1025  
Db 1553 -ProAsnGlyThrCysIleGlyPheGlnLeu-ValCysAspGlyGlnProAspCysGlyA 1572  
QY 1026 GGCCTGGCCCTGCTCCCTCA-----TGTAACACATATACATGCTTGG 1067  
Db 1572 rgProGlyGlnValGlyProSerProGluGluGlnGlyCysGly-----A 1587  
QY 1068 CACACACAGAGTG-----CACATGTCGCAAGACTCTCTCAGC 1106  
Db 1587 laTrpGlyProTrpSerProTrpGlyProCysSerArgThrCysGlyProTrpGlyGlnG 1607  
QY 1107 TGACACACAGATCCATTCTCAAGTATCTACTGATAGACACTCATGCTGCCAAGTCTCTCA 1166  
Db 1607 lyArgSerArg-----ArgCysSerProLeuG 1616  
QY 1167 TCCTCAACATACATGCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1207  
Db 1616 lyLeu-----LeuValLeuGlnAsnCysProGlyProGluHisGln 1630  
QY 1208 --GAGGAGTGTTCCT 1256  
Db 1630 erGlnAlaCysPheThrAlaAlaCysProValAspGlyGluTrpSer-ThrTrpSerPro 1649  
QY 1257 ---ACCCTCACCCCAAGAGTGGGACACACACCTGAGGGGTGCGAGCTG 1313  
Db 1650 TrpSerValCysSerGluProCysArgGlyThrMetThrArgGlnArgGlnCysHis--- 1668  
QY 1314 CTTCCCTGTGGC-----CGGCGCGGCTCATCTCTCTCTCTCTCTCTCTCTCTCT 1352  
Db 1669 SerProGlnAsnGlyGlyArgThrCysAlaAlaLeuProGlyGlyLeuHisSerThrArg 1688  
QY 1353 CCATCC-----TCCACACAGGGGACTCGTGGGATCTTAGCGTCCGTGACCGTGC 1406  
Db 1689 GlnThrIysProCysProGln-----AspGlyCys-----ProAsnAlaThrCys 1703  
QY 1407 TCTGGCACCCTCCCTG-----CAGTGCCAGCTCTCAATCGCAGGGGACC 1451  
Db 1704 SerGlyGluLeuMetPheGlnProCysAlaProCysProLeuThrCysAspAspIleSer 1723  
QY 1452 CAGGCGCTGGCCCTCCATCCAGAAACCTATGACTCACCCTGCTACCTGAGGACCAAC 1511  
Db 1724 GlyGlnValThrCysProAspTrpProCysGlySerProGlyCysTrp-CysProG 1743  
QY 1512 TCCGAGCTT-----GGCTGGACCTATGTCAGTATCCAGCTGAGGATCTGG 1559  
Db 1743 uGlyGlnValLeuGlySerGlyGlyTrp-----CysValTrpProArgGlnCysProC 1761  
QY 1560 GAGTTGGGAGAGTGAGGAGTGGGAAAGACACTCTCAACCTGAGGGT----- 1612  
Db 1761 sLeuValAspGlyAlaArgTyTrp-ProGlyGln-----ArgIleLysAlaA 1777  
QY 1613 -----CTGTAATATGATGGGTGAGAGG----- 1636  
Db 1777 spCysGlnLeuCysIleCysGlnAspGlyArgProArgArgCysArgLeuAsnProAspC 1797  
QY 1637 -----GGCTCTTGGCTCCACCTCCCTCTGCTGCTATCTCTCTCTCTCTCTCTCT 1691  
Db 1797 ysAlaGlyGluAlaLeuProSerGlySerLeuValLeuSerLeuAspArgProAlaA 1817  
QY 1692 TAGGTGGCCCCCCT 1748  
Db 1817 isProProProSerGlySerAspCysTrpProSerLeuSerGlyLeuValLeuVal 1837  
QY 1749 TCGCACCCGCTTGGC-----CCATTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1794  
Db 1837 euLeuValThrLeuGlyGlnValProGlyProLeuTrpIysProGluHisProValVal 1857  
QY 1795 -----CCGCGCT 1816  
Db 1857 euProGluLeuGlnProProProProLeuArgProArgSerProValProHisProP 1877  
QY 1817 GGAAGCTCTCTCTGTTTACACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1863  
Db 1877 roGlnGlyThrGlnThrGluProCysGlyGlyCysGluHisGlnGlyGlnValHisArg 1897  
QY 1864 --GGGTTGGGGGGGACAGAGGGGGGG-----A 1891  
Db 1897 alGlyGluArgTrp--HisGlyGlyProCysArgValCysGlnCysLeuHisAsnLeu 1916  
QY 1892 CTTCCCATCTCTG-----GTTCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1935  
Db 1916 hrAlaHisCysSerProTyTrpCysProLeuGlySerCysProGlnGlyTrpValLeuVal 1936

[illegible]



Db	2620	-----ProCysSerGlyGlyThrGlyLysGlyValLeuGlyTrpGlyHisGlyGly	2636
Qy	3261	TGTCAGGGACACAAAATAGAACTTTGGGAGCAGGTATCTCTTGGTGGTGAGCCAG	3320
Db	2637	-----SerThrValGlyThrGlyArgLeuGlyLeuProAlaPro	2649
Qy	3321	CGGCTC-----TGCCCTCTCTCT	3338
Db	2650	ArgLeuThrTrpCysProSerProThrArgLeuArgAlaGlyProCysValCysGluCys	2669
Qy	3339	-----TCCCATCACC-----TCCTCTTTT	3359
Db	2670	ArgSerValProGlyAlaGlyAlaProMetProThrLeuLeuProGlySerGingly	2689
Qy	3360	CACAGCTGAACCTCTGGCC-----CCCTTTTCA	3389
Db	2690	GlnGlnLysLeuGlnTrpAlaLeuCysGlySerLeuProSerLeuLeuCysProLeuGly	2709
Qy	3390	ACGAGCCAGACT---TCAACCTTCCCGCTGGGGCCAGAGACTCTGCCAGAGCCACTG	3446
Db	2710	LeuSerAlaLeuPheHisLeuPro-----GlyCysArgCysProProGlyLeuLeu	2727
Qy	3447	TTG---ACTTGAGGTGTGGCAGAGCTCAATGACAAACTCGGCTGACCCAGAACTACG	3503
Db	2728	LeuHisAspThrArgCysLeuProLeuSerGluCysProCysLeuValGlyGluLeu	2747
Qy	3504	AGG---CCTACAGCACCTTC-----TCGTGTACTTCGTGGCTCAACCGCTC	3548
Db	2748	LysTrpProGlyValSerPheLeuLeuGlyAsnSerGlnCysValCysGluLysGly	2767
Qy	3549	AGGCTG-----CCACTGTCTGAGCTGCGCGCAGAGCTTGGCCACT	3587
Db	2768	GluLeuLeuCysGlnProGlyGlyCysProLeuProCysGlyTrpSerAlaTrpSerSer	2787
Qy	3588	TCGTGACACCTCCAGGGCTCTGTGGCAGCAATTTCGGGGCGTCATGGCAGCTCTGGGCT	3647
Db	2788	TrpAlaProCysAspArgSerCysGlySerGlyValArgAlaArgPheArg-Ser-----	2805
Qy	3648	ACCACTGCCACCGCTGCTGGAGCTGAACCCACTTGGACTCTTGGCGCTGCCCA	3707
Db	2806	-ProSerAsnProProAlaAlaTrp-----GlyGlyAlaProCys-----	2818
Qy	3708	GTGACTTCTCTCCAGAGATGAGCAGCTTCTGCTGCTGAAGGA-----GC	3752
Db	2819	-----GluGlyAspArgGlnGluLeuGlnGlyCysHisThrValCysG1	2833
Qy	3753	TGCAGACTGGCTGTGGCTGTGGCAGAGACTTCAACGGCTCAAGAGAA-----	3804
Db	2833	yThrGlyIleAlaGlySerLeuGlyAlaGlyValProProSerSerGlnPheCysTh	2853
Qy	3804	-----	3804
Db	2853	rLeuArgThrHisGlyMetGlyProThrAspHisSerThrTrpGlyIleGluValPheG1	2873
Qy	3805	-----GATGACGCTCCAGCAGCTGCAGTCAACCTCGACCTCGACCTGG	3842
Db	2873	yTrpThrProTrpThrSerTrpSerSerCysSerGlnSerCysLeuAlaProGlyGlyG1	2893
Qy	3843	GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTTCGGCTCCCTTCAAAACCTTGCTCC	3902
Db	2893	yProGlyTrpArgSerArgSer-Arg-----	2901
Qy	3903	CACTTGTGAGAGCCGCTGTATGCCAACACCTGTTGACCCAGGAGACAGAGCTGTG	3962
Db	2902	--LeuCysProSer-----ProGlyAspAla-ThrGlnGluProCysSerProPro	2924
Qy	3963	AGCCTCTGGCCCTTCTCTGACCGGCTGGCGTGTGTCATGATCAGCCCTGTCTCTCTCCC	4022
Db	2912	-----ProGlyAspAla-ThrGlnGluProCysSerProPro	2924
Qy	4023	CACCTCCAAAGTCTACGAGCTGGGGAGGAGGTACAGTAGGCCCTGTCTGTCTCTGTT	4082

```

QY 4830 TACT-----CTCTTGGCCAGCCTCAC----- 4851
Db 3240 oGlyHisCysSerCysLeuAspLeuLeuThrGlyGlnArgHisProGlyAlaArgLe 3260
QY 4852 -----CCCTCTGGTGCTCAGACTCTTGGCCACCCCTGG----- 4887
Db 3260 uAlaArgProAspGlyCysAsnHisCysThrCysLeuGluGlyArgLeuAsnCysThrAs 3280
QY 4888 -CTCCCACTCCCTGCTGC-----CTCTG 4910
Db 3280 pleuProCysProAspCysGlyGlyGlnSerLeuHieProCysGlyGlnProCysPr 3300
QY 4911 GTGAGCTGCACAGAGCTCTGGGAAGAGGCCCTCTCTCCCTCCCGCA----- 4956
Db 3300 oArgSerCysGlnAspLeuSerProGlySerValCysGlnProGlySerValGlyCysG 3320
QY 4957 -----CTGGGGCGATGGCGGCACCTCAGACTT-----AC 4985
Db 3320 nProThrCysGlyCysProLeuGlyGlnLeuSerGlnAspGlyLeuCysValProAl 3340
QY 4986 CCACTGCTGCTGCACACCAACCC-----CTGATCCCTCAGTCTCCACACAGCTT 5039
Db 3340 aHisCysArgCysGlnTyrGlnProGlyAlaMetAlaProSerPheValPro----SerTh 3359
QY 5040 CTGT 5043
Db 3359 rCys 3360

RESULT 5
ID Q9SPM0 PRELIMINARY; PRT; 1315 AA.
AC Q9SPM0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Extensin-like protein.
GN PEX2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=POLLEN;
RX MEDLINE=21330248; PubMed=11437249;
RA Stratford S., Barnes W., Hohorst D.L., Sagert J.G., Cotter R.,
RA Golubiewski A., Showalter A.M., McCormick S., Bedinger P.;
RT "A leucine-rich repeat region is conserved in pollen extensin-like
RL Plant Mol. Biol. 46:43-56(2001).
DR EMBL; AF159297; AAD55980.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR002965; P_rich_extensin.
DR Pfam; PF00560; LRR; 3.
DR PRINTS; PR01217; PRICEXTENSIN.
DR SMART; SM00370; LRR; 3.
DR PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
SQ SEQUENCE 1315 AA; 134402 MW; 64C97A2A01F0936F CRC64;

```

```

Alignment Scores:
Pred. No.: 9,16e-20 Length: 1315
Score: 397.50 Matches: 301
Percent Similarity: 32.88% Conservative: 88
Best local Similarity: 25.44% Mismatches: 388
Query Match: 4.21% Indels: 407
DB: 10 Gaps: 52

```

US-09-931-704-3 (1-5087) x Q9SPM0 (1-1315)

```

QY 40 TCGCGGAGCGCGCTCGCCCTCCACTCCCGCAGCCTCCGGGAGGAGGAGCGCCGCCCC 99
Db 461 AlaProAlaProMetArgMetProThrLeuArgSerProAlaAspGluTyrIlePro 480
QY 100 GCCGGCCCGCCCGCCCGCCCTGAGCCTCCG-----AGCAGGTTGAAACCCCAACTAG 153
Db 481 ThrProValProAla-LysSerProGlyThrSerProProAlaSerArgGlyAl 500
QY 154 CCCTGCTCTTATACATGACAGCAGCGCCCGCCCTGATACCT---AAACCGACCAAGT 210
Db 500 aProLeuGlnAlaGlnProProAlaAlaSerPro-ProAlaThrProValLys 520
QY 211 CACAGCCCTCCA-----ACTCACCTCTGCCTGCCCGCAGACC-----TCACCA 252
Db 520 er-SerProProAlaAlaValValLeuProProAlaLysThrProSerProPro 539
QY 253 CATCTTTGTGACTCAAAACCT-----CAACCGCACTAAATCAACCAATCCCAAGTCTA 306
Db 540 AlaProValAlaSerProProGluAlaPro-ValSerSerProGlnProGlnVal-- 558
QY 307 AACTAATCTGAACCTTTTAAAGTAACCCAGCTCTTAAACCTAACCTAGCCCAATGCAAT 366
Db 559 -----LysSerPro-----ProProAlaProva 567
QY 367 TATATCTACCCCTAGCCAAACCCCTAACTGCCTTTGGCAGTCCAAAGTGTCCACTGAATCCT 426
Db 567 lAlaSerProProProMetLysSerProProProProAlaArgVal-AlaSerProP 587
QY 427 CACCTTGGTCTCTACTGAAATCCAGAAAGCATATTTCCCACTGCCACATCCCTCC 486
Db 587 roProLeuMetLysSerProProProAlaProValAlaSerProProGlnProLeuL 607
QY 487 TTACAGACCCCAACCCCTGCTCG-----ACTCCTGGTATCTCTGGGATGTGCCAAACT 540
Db 607 ysSerProProProValLeuThrLeuSerThrProSerValLysSer---ProProP 626
QY 541 CTGAGTGCCCATCAGCAACCAAGCCCGACTCGTCAAAAT-----GCACCTC 585
Db 626 roProValProValAlaSerProProProProValLysSerProProLeuAlaProV 646
QY 586 TCTCCTTCCTGTCTC-----CCACCTTGCAGGCTGATGAAGGCTCA 630
Db 646 alSerSerPro-SerProProValLysLeuProProLeuProAlaProGlyLysSerThr 665
QY 631 TTGAAGTCCAACCTTTTCCCACTTAACACCAAGAGGGGTGAACCTCCACACTGCCACC 690
Db 666 ProProProGluGluGluLysPro-ThrProThrProValLysSerSerProProPr 685
QY 691 GTTCCCTGAGAGTGAGCACTAAATCTCCTTCAATCAACCCCACTACACTTCCACAC 750
Db 685 o-----GluLysSerLeuProProProThrLeuThrSe 697
QY 751 TCAGGAATCACATCTCTAGATATACCCAAAATAAGCCCCATAAGCAGCCCGACCTAG 810
Db 697 rProProProGlnGluLysProThrProSerThrProSerLysProProProSe 717
QY 811 TGGTCTAACCTTATACCTTGTCTCTATGGGTGAGTCTGTCTTCTGGCGCGCTCTCTC 870
Db 717 rPro-----ValGluThrLeuProProProSerLy 727
QY 871 CTGCTCTCCTTAGAGCTGACTGTGTCTAGCTGCCAGCTGCGAGCTGTGACATGTGTCTCC 930
Db 727 sSerSerProProGluGluProValSerSerProProGlnAlaProLysSerSerPr 747
QY 931 CACCTCTGACTCCCTCAAGCTGAGTGGGAGCTGGAAGACTGGCAGGAGCTAGGTAC 990
Db 747 o-----ProAlaProValSerSerProProProLeuLysSerSer----- 760
QY 991 AACTGGAACACAGCGAGGTGACCTGAGCTCCCTAGGCTGGCCCGCTCCCTCCATGTAC 1050
Db 761 -----ProProProValProGluSerSerProProPro----- 771
QY 1051 ACATATATATGTTGGCACACACAGTGGCACACATGCGCAAGACTCTCTAGCTGAC 1110

```

```
Db 772 -----Th- ProLysSerSerProProLeu 779
Qy 1111 ACACAGATCCATTCTCAAGTATCTACTATAGACACTCATCGTGCAGAGTCTCATCT 1170
Db 780 AlaProValSerSerProProGlnValGluLysThrSerProProAlaPro----- 797
Qy 1171 CAACATACACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1230
Db 798 -----ValSerSerProProProThrProLysSerSerProProLeuAla 812
Qy 1231 CCCCTCTGCCCTCCATCTGGTGTGTC-----CACACC-----TCA 1263
Db 813 ProValSerSerProProGlnValGluLysThrSerProProProAlaProValSerSer 832
Qy 1264 CCCCCCAGCCAGCCAAAGTGGGACACACACTGAGGGGTGCCAGCTGTCTCCCGCTG 1323
Db 833 -ProProProThrProLys-----SerSerProProLeuAlaProValSerSerPro-- 849
Qy 1324 TGGGCCCGGGCGGCTCATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1383
Db 849 ----- 849
Qy 1384 GTTAGCGTCTGTCACGGTGTCTGTGGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1443
Db 850 -----ProGlnValGluLysTh 855
Qy 1444 AGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCGCTACTCGGA 1503
Db 855 rSerProPro--Pro-AlaProValSerSerProProLeu--GluProLysProSerS 873
Qy 1504 GCACCACTCCGCACTTGGCTGGGACCTATGTAGTATCCAGCGTAGGAATCTGGGAGT 1563
Db 873 erProProSerValSerSerProProThrVal----- 885
Qy 1564 TGGGAGGAGTGAGGAGTTGGGGAAGACAGTCTTAACCGTGGAGGGTCTGGTAATGA 1623
Db 886 -----LysSerSerPro----- 889
Qy 1624 TGGGTGAGGAGGGCTCTTTGGCTCCACAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 1683
Db 890 -----ProProAlaProLeuSerSer--ProProMet 900
Qy 1684 TTCCCTCTTAGTGGCCGCCCTCTCCCATCTCCGCTGGCCAGGACTAGGCATGTGGGC 1743
Db 900 hrProLysSerSerProProAlaHisValSerSerProProGlu-----AlaG 917
Qy 1744 AGGCTCGACCCGCTTGGCCCATTTGCCATTTGCCATCTGGCTGCCAGCCAGCCGCGCTCC 1803
Db 917 luLysSerSerPro-----ProLeuAlaProIle-----Ser-SerProProSerGlu 932
Qy 1804 CCTGGGGCCGGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1863
Db 933 ProLysSerSerSerProProMet-ValGluLysThrSerProProProAlaThrVa 952
Qy 1864 GGGGTGGGTGGGACAGAGGGGGCCCACTCCCATCTCTGGCTTCCAGCTCGCCTCTGC 1923
Db 952 1-----SerSerProProProThrPro-LysSerSerProProProA 966
Qy 1924 CCCCAGACTGGGGCCCTGCTGCTGTGGACCCAGGGGCTCTCCCTCCGCTGCTCTCTCC 1983
Db 966 laProValSerSerPro-----ProProValValLysSerS 978
Qy 1984 ATCTAGCTGGGCCCTCTTAGGGGGTCAATGGGGGAAGGGAGCTGTAGGGAACCCAGG 2043
Db 978 erPro----- 979
Qy 2044 TAGTGGCAGGGGTTTAGGGTGTGGATGGAGGTTATGCTGTAAAGATTGGGGGTGGTCC 2103
Db 979 ----- 979
Qy 2104 AGAGGTGTTACAGAGCCCGAGGAGAGAGGAAGGAGGGTTGGAGGAGCGGACCACTG 2163
Db -----
Db 980 -----ProProAlaProValS 985
Qy 2164 GGGAAACGGCCCTCTCTCGTGTCTCTCTTCCACATCCACAGCCCTACTCTGTGAGCCA 2223
Db 985 erSerProProProThrProLys--ProLeuProProProAlaProValSerSer---- 1002
Qy 2224 GGGAAAGAAAGGAAGAGTGGGGGGAGCTGGTCCAGCCCGCAGGATACACCGAGG 2283
Db 1003 -----ProProProValValLysSerSerProProProThrPro----- 1015
Qy 2284 AAATTAGTTGTCTCTGTGCTTGTGACGGTGTGAACCTCCCTCGGGCCCTTGCCTATCC 2343
Db 1016 -----ValSerSerProProProProThrProLysProLeuP 1027
Qy 2344 CAGGCTCTCCCTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2403
Db 1027 ro--ProProProThrProValSerSerProProThr**LysProLeuProProPr 1046
Qy 2404 CTGGGCCCCAGCGCTCCCGAGGGTTGGAAGGGCTCTGCCCTCTCTCTCTCTCTCTCT 2463
Db 1046 oAlaProValSerSerProProProValValLysSerSerProPro--ProAlaProVa 1065
Qy 2464 CTGCTTCTCCATAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2518
Db 1065 lSerLeuProProProThrArgLysProSerProProProArgThrArgValSerSerProAr 1085
Qy 2519 -----TCTGCAACCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2568
Db 1085 gProValValLysCysCysProProPro-----ThrLeuValSer-SerProProP 1102
Qy 2569 AGTACCG--TCTCTCCCGAGCCCTCAGCTGGTGGCGCTGGGTGTCTACCGGCAAT 2625
Db 1102 roAlaProLysSerLeuProProThr----- 1111
Qy 2626 GGGGTCTGGTTCCAAATGGGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2685
Db 1112 -----ProValSerSerProProProGluValLysSerSerPro- 1124
Qy 2686 TGCTTCACTCCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2745
Db 1125 -----ProProThrProValSerSerProProPro----- 1134
Qy 2746 AAATTTCTCAAGAGTGGTGTACACCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2805
Db 1135 -----AlaProLysSer--SerProProProThrProValSerSerProProGlu 1152
Qy 2806 CTTAAC-----CCCTGCAATCTGGCTTCCAGGCCCGCAGCAATGTTCTCTCCAAG 2856
Db 1152 uLysSerSerProProProAlaProValSerSerProProSer-----AlaProLy 1169
Qy 2857 GTGTCAGGCACCTCTCTTGGCAAGCCGACAGTGTGTTGAAGGCTCATCTCTCTCTCT 2916
Db 1169 sSerSer--ProProProAlaPro-----ValS 1178
Qy 2917 CTGTTTGGAGCCACACTGTGCTGAGCGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2976
Db 1178 erLeuProProProGluVal-----LysSerSerProProProAlaProI 1193
Qy 2977 TGCATCTCTCTGGGCGACCTTCTACTCTCTCAGCTCTCTCTCTCTCTCTCTCTCTCT 3036
Db 1193 leSerSerProProProAlaLysSerProProProProAlaProMetSerSerLeu- 1212
Qy 3037 CCTGCCCCCAGAGCGGCACTCTCCCAAGTTTGGCCACCCAGCCCAATCAGCAGCTCT 3096
Db 1213 -----ProProProValLysSerProProP 1221
Qy 3097 CTGAGCGTGTGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3146
Db 1221 roProAlaProValSerSerProProProProMetLysSerProProProAlaProI 1241
Qy 3147 TGGAGAGCTCACCGGCACTGTCTCACTGTCCACTGTCCACTGATCAATATCTCTCTAT 3206
Db 1241 le---SerSerProProPro----- 1246
```

```
QY 3207 GGAAGAACTCAGGAGGCCATGAACAAAGAGCCCTAGCATGGAGACAGGGCCAGTGTGAG 3266
Db 1247 -----AlaProVal----- 1249
QY 3267 GGGACACAAAATAGAAACTTTGGGAGCAGGTATCTCTTGGTGGAGCCAGCGGCTC 3326
Db 1250 -----LysProProSerL 1254
QY 3327 TGCCCTCTCTCCCATCCCTCTCCCTTTTACAGCTGAACCTAGCTGGGCCCCCTT 3386
Db 1254 eu-ProProAlaProAlaProValSerProProAla---ValThrSerAlaProPro 1272
QY 3387 TCAACGAGCCAGACTTCAACCCCTCC 3413
Db 1273 LysLysGluGluAapSerThrAlaPro 1281

RESULT 6
Q9XDH2 PRELIMINARY; PRT; 763 AA.
AC Q9XDH2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Proline-rich mucin homolog.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=20090472; PubMed=10627046;
RA Espitia C., Lacleite J.P., Mondragon-Palomino M., Amador A.,
RA Campuzano J., Martens A., Singh M., Cicero R., Zhang Y., Moreno C.;
RT "The PE-PGRS glycine-rich proteins of Mycobacterium tuberculosis: a
RT new family of fibronectin-binding proteins?";
RL Microbiology 145:3487-3495(1999).
DR EMBL; AF071081; AAD41594.1; -.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistol_extensin.
DR InterPro; IPR002965; P_rich_extensin.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01582; KV33CHANNEL.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSTLEXTENSIN.
SQ SEQUENCE 763 AA; 75035 MW; 39168EC45A5916F8 CRC64;

Alignment Scores:
Pred. No.: 9,51e-20 Length: 763
Score: 396.50 Matches: 291
Percent Similarity: 32.86% Conservative: 77
Best Local Similarity: 25.98% Mismatches: 313
Query Match: 4.20% Indels: 441
DB: 2 Gaps: 59

US-09-931-704-3 (1-5087) x Q9XDH2 (1-763)
QY 1209 AGGAGTGTTCCTCCCT-----CCTCCATCCCTCTGCTCCCTCCCTGCTGTCCTCC 1256
Db 8 ArgAlaLeuAlaProLeuProProAlaProProAlaProAlaProAlaGlu-ProLysSerLysPr 27
QY 1257 ACCTCACCACCCACCCAGCCAGCCAGGTGGGGAGCAGACACCTGAGGGGCTGCCAGCTGCT 1316
Db 27 oProPheProProAlaProProAlaPro----- 36
QY 1317 CCCCGTGTGGCGCGCGCGCTCATGCTTCGTCCTCCCTGCTCCACAGGGGACTCGT 1376
Db . 37 -ProCysftrp-----MetLeuVal----- 42
QY 1377 GGGGGATGTTAGGCTGTGACGCGTCTGTCACGCTGCTGGGACCTCCCTGACAGTGCACGCTCTCA 1436
Db 43 -----SerAlaAlaProProCysProProAlaProPr 53
```

```
QY 1437 ATGCACAGGGGACCCAGGGCTGG-----CCCTCCATCCAGA 1475
Db 53 oAlaProProLysProLysSerLysAlaProPheProProAlaProAlaProAla 73
QY 1476 AAACCTATGACCTACCCGCTACCTGGAGCACCAACTCCGAGCTTGGCTGGGACCTATG 1535
Db 73 aArgGluLeuAlaProProLeuProProAlaProPro----- 85
QY 1536 TGAGTATCCAGGTAGGAATCTGGGAGTGGGGAGGAGTGAGGATGGGAAAGACAGT 1595
Db 86 -GluAlaPro-ArgGluSer----- 91
QY 1596 CCTAACCGTGGAGGGTCTTGTTAAATGATGGGGTGAGGAGGGCTCTTTGGCTCCACCA 1655
Db 92 -----ArgProAlaLeuProProCysp 99
QY 1656 GTCCCTCTGTGTCTATCTCTCTCCCTCTCTTAGTGGCCCCCCCCACTTCCCAT 1715
Db 99 roProProValValIleProAspPro-ProGluProAlaAlaProProValPro--- 117
QY 1716 CCCTGGCCCCAGGACTAGGCATGTGGCAGGCTCGCACCCGCTTGGGCCATGGCCCA 1775
Db 118 ProAlaPro-----AsnSerProProPheProProPhePro 129
QY 1776 CTGGCTGCCAGCCAGCCCGCTCCCTGGGGCGCG---GGGAAGTCTCTCTGTT 1832
Db 130 ProAla-ProLysPheValProAlaProProValProProValProAsnSerProProPh 149
QY 1833 TACACCGTGTGTGTCTCTTCTGCGGGGGGGTGTGGGTGGGGACAGAGGGGCCCCAC 1892
Db 149 eProPro----- 151
QY 1893 CTCCATCCCTCGGTTCAGCTCGCTCTGCCCCCAGACCTGGGGCCCTCTCTCTGGA 1952
Db 152 -----PheProAlaAlaLeuAs 158
QY 1953 CCCAGGGGCTCCCTTCCGCTCTGCTCTCCATCTAGTGGGGCTCTTAGGGGGTCTCAT 2012
Db 158 nProProAla-----ProProAlaProProLeuAlaAsnSerProProLeuProAl 176
QY 2013 GGGGAAGGGGACTGTAGGGAACCCAGCAGTAGTGGCAGGGGGTTTAGGGTGTGATGG 2072
Db 176 a-ProProThrProAlaGlyThrProProAlaAla-----ProTrp----- 189
QY 2073 AGGTTATGCTTAAGGATTTGGGGTGTGTCCAGAGTGTTCAGAGAGCCAGAGAGAG 2132
Db 190 -----ProProVal-ProAlaAlaProLysSerLys 199
QY 2133 GAAGGAGGTTTGGAGGCCAGGACCCATGGGGAACCGGCCCTCTTCCCGTGT--- 2189
Db 200 ProAla-----SerProProArgProProAlaProProMetProAlaThrProMetGlu 217
QY 2190 -----CCTCTTCCATCCAGACCCCTACTCTGGAGCCAGGAAAGAAAGGAGAGAG 2243
Db 218 PheProProLeuProProValProPro-----AspPro 228
QY 2244 GTGGCGGGGAGCTGGCTCCAGCCCGCAGATACACCGAGGAATTAGTTGTCTCTGTGC 2303
Db 229 IleSerLysGluThrProProAlaPro----- 237
QY 2304 TTGTCAGCGTGTGAACCTCCCGCTTGGGGCTTGGCTATCCAGGCGCTCTCCCTCTCTTC 2363
Db 238 --ProAlaProProIleProProAlaProValProIlePro--ProValProProLeuP 256
QY 2364 TCCCTTCTTCCAGTTATACATCTCCCTCATCTTCCCTGGGGCCCCAGCGGCTCCCC 2423
Db 256 roPro-----ValProAsnLysIleProProAlaProProAlaProP 270
QY 2424 CGAGGGTTGGAAGGGCTCTGCCCTCTTCCCTATACCATGCTGTCTTCCATAGCCTTCT 2483
Db 270 roValAla-----ValAlaAlaValLeuValAlaProCys-----Pro-ProLeuPro 285
```





Db 358 -----IleGlnGlyHisProProGlySerLeuTyProProAsnSerValLeu 374  
Qy 1025 AGGCTGGCCCGCTCCCTCCATGATACACATATACATGTTGGACACACACAGTGGCAC 1084  
Db 375 AlaProGlySerIleProPro--LeuArgProSerAsnGlnGlyHisProProGlys 394  
Qy 1085 AC---ATGCCAAGAGCTCTCTCAGCTGACACACACATCCATCTCAAGTATCTACTGATA 1141  
Db 394 erLeuLeuProProAsnArgAlaLeuPheArgLeuLeuAspHisGlnIleLeuAspA 414  
Qy 1142 GACACTATGCTGCGCAAGTCTCATCTCATCTCAACATACATAGCTCTCTCTCTCCCG 1201  
Db 414 snHis-----GlnAspHisCysLeuLeuMetTrpGluT 426  
Qy 1202 TCTTGCAGGAGTGTTCCTCCCTCCATCCCTCTGCTCCCATCTCCATCTGGTGTCCACCT 1261  
Db 426 yrIlePro-----HisPro-----ArgAlaGlyPheProProS 437  
Qy 1262 CACCCCCACCCAGCCCAAGTGGGGACACACACCTGAGGGCTGCCAGCTGCTTCCCG 1321  
Db 437 erGlyProLeu--ThrGlnGlyGly-----HisProThrGlySerLeuLeuProProAs 454  
Qy 1322 TGTGGGC-----CGGGCGCGCTCATGCTTCTCGTCACTCTGCTCCACAGGGACTCG 1375  
Db 454 nThrGlyLeuProProGlySerIleProProLeuArgProPro----- 468  
Qy 1376 TGGGGATGTAGCTGCTGTGCAGGTGTCTGGCACCTCCCTGGCAGTGCACCTCTC 1435  
Db 468 ----- 468  
Qy 1436 AATCGACAGGGACCCAGGGCTGGCCCTCCCA---TCCAGAAACCTATGACCTCAC 1492  
Db 469 -----AsnGlnGlyGlyHisProProGlySerGlnLysPro----- 480  
Qy 1493 CGCTACTGGAGCAACAACCTCCGAGCTTGGCTGGGACCTATGTGAGTATCCAGCGTAGG 1552  
Db 481 -----ProAsnValG 484  
Qy 1553 AATCTGGGAGTTGGGGAGGTGAGGAGTTGGGAAAGACAGTCTCAACGTTGGAGGTT 1612  
Db 484 ylleTy-----ProProSerThrGly----- 491  
Qy 1613 CTGGTAAATGATGGGTGAGGAGGGCTCTTTGGCTCCACCCAGTCTC----- 1658  
Db 492 -----TrpIleProProPheGlyProLeuThrG 501  
Qy 1659 -----CCCTCTCTGGTCTATCTCT-----GCCCTCTCTCTTAGTGGC 1699  
Db 501 nGlyGlyHisProProGlySerLeuLeuProProAsnThrGlyLeuProProGlySerI 521  
Qy 1700 CCCCCCATCTCCCTCCCTGGCCCGGAGTGGGAGTGGGAGGCTCGCACCCGCG 1759  
Db 521 eProProLeuGlySerPro-----AsnGlnIleGlyHisProPr 534  
Qy 1760 TTGGC---CCATTTGCCCACTGGCTGCCAGCCCGCCGCTCCCTCCCTGGGGCGCG 1816  
Db 534 o-GlySerGlnLysProProAsnValGlyIleTyProProSerThrGlyTrpThrProp 554  
Qy 1817 GGAAGTCTCTCTGTTTACACCGTGTGTGGTGTCTCTTGGCGGGGGTGGTGGG 1876  
Db 554 roSerGlyProLeu----- 558  
Qy 1877 GACAGAGGGGCCCCACCTCCCATGCTGCTGCTCCAGCTCGCTCTGCCCCAGACCTGGG 1936  
Db 559 -----ThrGlnGlyGlyHisProProGlySerLeuLeuProProAsnThrG 574  
Qy 1937 GCCCTGCTCTGAGACCGAGGGCTCCCTTCCTCGCTCGCTCTCCCATCTCTAGCTGGG 1996  
Db 574 ly-----LeuProProGlySerIleProProLeuGlySerProAsnGlnIleGlyH 591  
Qy 1997 CTCCTAGGGGGTCTAGGGGAGGGGAGTCTAGGG---AACCCAGGAGTGTAGTGCAGG 2053  
Db 591 isPro-ProGlySerGlnLysProProAsnValGlyIleTyProProSerThrGly--- 609

Qy 2054 GGGTTTAGGGTGTGATGAGGTTATGCTGTAGGATTGGGGTGGTCCAGAGGTGTC 2113  
Db 610 -----TrpIlePro-----ProSerGlyPro-----Leu 617  
Qy 2114 AGAGAGCCAGAGAGAGAGAGGAGGTGTGAGAGCGGAGGACCAATGGGAAACCGGC 2173  
Db 618 ThrGln-SerGlyHisProProGlySerLeuLeuProPro--AanThrGlyLeuPro-- 635  
Qy 2174 CCCCTCTCTCCGTTCTCTCTTCCACATCCAGAGCCCTACTCTGGAGCCAGGAAAGAA 2233  
Db 636 -ProGlySerIleProProLeuArgProProAsnGlnGlySer----- 649  
Qy 2234 AGGGAAGAAGTGGCGGGAGCTGGCTCCAGCCCGCCAGATACACCGAGGAAATAGTTT 2293  
Db 650 -----LeuLeuProProAsnThrGlyLeu----- 657  
Qy 2294 GTCTCTGTGTGTGACGCTGTGAACCTCCCTGGGCGCTTGCCTATATCCAGGCTCTC 2353  
Db 658 -----ProSerGlySerIlePro----- 663  
Qy 2354 CCCTGTCTCTCCCTTCTTCCAGTTATACATCTCCCTCATCTCCCTTCCCTGGGCCCA 2413  
Db 664 -----ProLeuGlySerProAs 669  
Qy 2414 GCCGCTC-----CCCGAGGTTGAAAGGCTCTGCCCTCTCCCTATATACATGCTGT 2467  
Db 669 nGlnIleGlyHisProProGlySerGlnLysProProAsnSerGlyIleTyProProSe 689  
Qy 2468 CTTCCATAGCTTCTCT-----CCTGTCTCTATCATAGACTGCTTCCA-----TT 2512  
Db 689 rThrGlySerIleProProSerGlyProLeuThrGlnGlyHisProProGlySerLe 709  
Qy 2513 TCTTCTTCTGCAACCTCTCTCATCTCTCTCTTCTTTCGAGTGTAGTCAAGTA 2572  
Db 709 uLeuProSerAsnThrGlyLeu----- 716  
Qy 2573 CCGCTCTCTCCCGACCTCTCAGCTGGTGGGCTGGTGTGTGACGGCAATGGGGCTC 2632  
Db 717 -ProSerGlySerIleProProLeuGlySerProAsnGlnIleGlyHisProProGlySe 736  
Qy 2633 TGGTTCCAAATGGGCGACCTCTCATCTCTCTCTTGTCTTGTGCGAGAAACCTTGTCTCA 2692  
Db 736 r-----GlnLysPro----- 739  
Qy 2693 CTCCTACTGCCCTCTCTAGTTCGCGACCTTTTCTCTCTCTGCTTTCCTGCAAAATTC 2752  
Db 739 ----- 739  
Qy 2753 TCCAAGAGTGTGTACACCTCTCTGCTCTCACTCTCTCTCCACCCACTCACTTCTTAACC 2812  
Db 740 -ProAsnSerGlyIleTyPro-----Pr 747  
Qy 2813 CCTCAATCTGGCTTCAGGCGCCAGCAGCAATGGTCTCTCAAGGTGCTGAGCAGCTCC 2872  
Db 747 oSerThrGlyTrpIlePro--ProSerGlyProLeuThrGlnGlyHisProPro-- 765  
Qy 2873 TTGCCAAGCCCGACAGTGTGTTGAAGGCTCATCTCTTGTCTGTCTTGTGAGCCACA 2932  
Db 766 -----GlySer----- 767  
Qy 2933 CTGCTGAGCGCTGCTGCTCTCTCGAACTCTCTCTCTGCTGCTCTGCTCTCTCTCT 2989  
Db 768 -----LeuLeuProProAsnThrGlyLeuProProGlySerIleSerProLeuG 784  
Qy 2990 GCCACTTCTACTCTCCAGCTCTCCAGGCTCTCTTCC-----TCTCTGTC 3037  
Db 784 ySerProAsnGlnIleGlyHisProProGlySerGlnLysProProAsnValGlyIleTy 804  
Qy 3038 CTGCCCCCAGCGGGGACCTCTCCCAAGGTTTGGCCACCCAGCAGCAATCAGCAGCTCTTC 3097  
Db 804 rProProSerThrGlySerIleProProLeuArgProProAsnGlnGlyHisPro-- 823





Db 602 ProLysCysTyrProGlySerSerAspProGluCysLeuAsnCysTyrProGlySerPro 621  
:::  
Qy 142 AACCCAACTAGCC-----TGCTCTTCATACATAC 174  
:::|||||  
Db 622 AspProArgCysProLysValProThrThrLysLysAlaGlyCysTyrGluGlySerAsp 641  
:::|||||  
Qy 175 AAGCAGCC-----CCCATCTGATACCTAAACCCAGCAAGTCACAGCCCTCCAACT--- 225  
:::|||||  
Db 642 AspProArgCysGlnProAlaThrThrLysLysProSerThrArgArgProProThrIle 661  
:::|||||  
Qy 226 CACCTCTGCCT-----GCCAGACCTCACACATCTCTGTGGACTCAAACT 273  
:::|||||  
Db 662 AlaProArgProArgCysTyrLeuGlySerThrAspPro---SerCysProGlnGlnThr 680  
:::|||||  
Qy 274 CAACCGCACTAAATCAACCAATC-----CCAAGTCTAAAC 309  
:::|||||  
Db 681 GlnProThrThrArgThrProIleThrThrLysLysProArgCysPheProGlySerAsn 700  
:::|||||  
Qy 310 TAATCTGAACCTTTAAAGTAACCCAGTCCTTAAACCTAAACCTAGCCCAATGCCAATTAT 369  
:::|||||  
Db 701 AspProGluCysGlnProAlaThrThrArgArgProLeuThrThrSerLysProAsn--- 719  
:::|||||  
Qy 370 ATCTACCTAGCAAACTCACTGCTTTCGCACTTCCAAAGTGTCCACTGAATCTCTCAC 429  
:::|||||  
Db 720 CysTyrProGlySerThr-----AspArgArgCysProGlnGlnProVal 734  
:::|||||  
Qy 430 CTGTGCTCTCACTGAATATCCAGAAAGCATATTTCCCACTGCCCACATCCCTCTTA 489  
:::|||||  
Db 735 Thr-----ThrSerArgProArgCysTyrProGlySerSerAspProGluCys 750  
:::|||||  
Qy 490 CAGCAACCAACCTGGCTCTGACTCTGCTGATCTGGATGTCCAACTCTGCAG--- 546  
:::|||||  
Db 751 GlnProAlaThrLeuValProProThrThrValArgThrIleThrThrSerArgGln 770  
:::|||||  
Qy 547 -----TGCCATCAGCCAAAGCCGAGCTCGTCAA 576  
:::|||||  
Db 771 ArgCysTyrProGlySerSerAspProSerCysGlnProProThrThrArgLysProIle 790  
:::|||||  
Qy 577 ATGCACTCTCTCCCTTC-----TGTCCTCCCACTTCG 609  
:::|||||  
Db 790 eThrThrProLysProAsnCysTyrProGlySerThrAspSerArgCysProGlnThrPr 810  
:::|||||  
Qy 610 CAGGCTGATGGAAGGCTCATTAAGTCCA-----ACTTTTCCCACTAACACCAAG 663  
:::|||||  
Db 810 oArgThrThrGlnGlnProArgCysTyrProGlySerThrAspProAspCysGlnProAl 830  
:::|||||  
Qy 664 AACGGGTGAACCTCCACACTGCCCGTTCCTCGAGAGTGAACATAATCTCTCTCAA 723  
:::|||||  
Db 830 aThr----- 831  
:::|||||  
Qy 724 TCTAACCCACCTACACTTCCACACT-----CAGGAATCACATCTAGATATACCC 777  
:::|||||  
Db 832 -TyrLeuProProThrThrAlaArgThrThrIleThrThrAlaLysProArgCysTyrPr 851  
:::|||||  
Qy 778 AAAACTAAGC---CCATAAGGAGCCGACCTAGTGCTA---ACCTATACCTTGC 831  
:::|||||  
Db 851 oGlySerSerAspProTyrCysGlnProProThrThrLysLysProAsnCysTyrProGl 871  
:::|||||  
Qy 832 TTCCTATGGGTGAGTGTCTTGGGGCGGCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCT 891  
:::|||||  
Db 871 ySerThrAspSerArgCysProGlnLysProIleThr-----Pro-LysProL 888  
:::|||||  
Qy 892 ACTGTGCTCAGCTGCAGCTCTGACATGCTGCTCTC-----CCACCTCTGACT 942  
:::|||||  
Db 888 yCys---TyrProGlySerThrAspLeuGluCysLeuAsnCysTyrProGlySerProA 907  
:::|||||  
Qy 943 CCCCTCAAGCTGAGTGGAGCTGGAGAGCTGGCAGGAAGTGGGTACAACTGGGAACA 1002  
:::|||||  
Db 907 spProArgCysPro---LysVal-ProThrThrLysLysAlaGlyCysPheAspGlySer 925  
:::|||||  
Qy 1003 GGCAGTGCAGCTCAGTCCCTAGGCTGGCCCGCTCTCCATGTACACATATACAT 1062  
:::|||||

Db 926 AspAspProLysCysGlnProAlaThrThrLysLeuProProSerThrLysArg----- 942  
:::|||||  
Qy 1063 GTTGGCACACACAGTGGGCACACATGCCAAAGACTCTCTCAGCTGCACACAG----- 1116  
:::|||||  
Db 943 -----ThrProThrIleAlaProLysProArgCysTyrProGlySerLysAspProGly 960  
:::|||||  
Qy 1117 -----ATCCATTCTCAAGTATCTACTATGATGACACTCATGGTGCCA--- 1158  
:::|||||  
Db 961 CysProGlnIleThrThrArgAlaProThrSerThrSerArgProArgCysTyrProGly 980  
:::|||||  
Qy 1159 AGTCTCATCTCTCAAAACATACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1215  
:::|||||  
Db 981 SerThrAspProGluCysHisProThrThrSer---SerProAlaIleThrArgIlePro 999  
:::|||||  
Qy 1216 -----TTTCCCTC-----CTCCATCCCTCTGCTCTCTCTCTCTCTCTCTCTCT 1257  
:::|||||  
Db 1000 ValThrThrArgIleProLeuThrThrAlaLysProArgCysTyrProGlySerGlnGlu 1019  
:::|||||  
Qy 1258 CCC-----TCACCCCCACCCAGCCCAAGGTGGGAGACAGACCTCTGAGGGCTGCCAGC 1311  
:::|||||  
Db 1020 ProGlyCysGlnProAlaThrArgProIleThrThr-----SerLysProArg 1036  
:::|||||  
Qy 1312 TGCTTCCCGTGTGGGCGGCGGCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1371  
:::|||||  
Db 1037 CysTyrPro----- 1039  
:::|||||  
Qy 1372 CTCGTGGGGATGTTAGCGTGCCTGTGCACCGTGTCTGGCACCTCTCTCTCTCTCTCTCTCT 1431  
:::|||||  
Db 1040 -----GlySerLeuAspProGluCysGlnProSer 1049  
:::|||||  
Qy 1432 TCTCAATCGCACAGGGGACCCAGGCGCTGGCGCTTCCATCCAGAAACCTATGACCTCAC 1491  
:::|||||  
Db 1050 ThrTyr----- 1051  
:::|||||  
Qy 1492 CCGCTACTCTGGAGCACCACTCCGAGCTTGGCTGGACCTATGTAGTATCCAGCGTAG 1551  
:::|||||  
Db 1052 ---LeuPro-----ProThrPro----- 1056  
:::|||||  
Qy 1552 GAATCTGGAGTGGGAGGAGTGAAGTGGGAGAAAGACAGTCTTAACCGTGGAGGT 1611  
:::|||||  
Db 1057 -----ValArgThrThrValProThrIleProThrThr-ArgIle 1069  
:::|||||  
Qy 1612 TCTGTAAATGATGGGTGAGGAGGCTCTTTGGCTCCACAGTCCCTCTCTCTCTCTCTCTCTCT 1671  
:::|||||  
Db 1069 eProValThrThr-----SerLysProAsnCysTyrProGlySerThrAs 1084  
:::|||||  
Qy 1672 TATCT 1725  
:::|||||  
Db 1084 pArgArgCysProLys-GluProValThrThrProLysProArgCysTyrProGlySerP 1104  
:::|||||  
Qy 1726 AGGACTAGGCACTGGGAGGCTCGCACCCGCTTGGCCCATTCGCCACTGGCTGCCA 1785  
:::|||||  
Db 1104 roAsnProGluCysGlnLysAlaThrThrThrSerPro----- 1115  
:::|||||  
Qy 1786 GCCAGCCGCT 1845  
:::|||||  
Db 1116 --ProThrThrArgThrProValThrThrSerLysProAsn----- 1128  
:::|||||  
Qy 1846 GGTGTCT 1905  
:::|||||  
Db 1129 --CysTyr-----ProGlySerThrAspSerArgCysProGlnLysPro----- 1142  
:::|||||  
Qy 1906 GTTCCAGCTCGCTCTGCCCCCAGACCTGGGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1965  
:::|||||  
Db 1143 -----ProThrThrLeuLysProLysCysSer---ProGlySerSerA 1156  
:::|||||  
Qy 1966 CTTCCGTCT 2025  
:::|||||  
Db 1156 spProGluCysLeuAsn----- 1161  
:::|||||  
Qy 2026 TGTAGGAACCCAGGAGTAGTGGCAGGGGTTTAGGGTGTGTAGTGGAGGTTATGTGTATA 2085  
:::|||||  
Db 1162 -----CysT 1163  
:::|||||

QY 2086 AGGATTGGGGTGTCCAGAGGTGTTCCAGAGCCAGGAGAGAGAGAGGGTTG- 2144  
Db 1163 yProGlySerProAspProArgCysProLysValProThrThrLysLysSerGlyCysP 1193  
QY 2145 -----GAGGAGCCAGGACCATGGGAAACCGCC-----CCCTCTT 2181  
Db 1183 heAspGlySerGlnAspProArg-----CysGlnProAlaThrTyLeuProProSerS 1201  
QY 2182 CCGCTGTTCT-----CTTCCACATCCAGACCCCTACTCTGGAGCCAGGAGAAAGAA 2232  
Db 1201 erArgArgProProThrThrAlaProLysProArgCysHisProGlySerThrAspProS 1221  
QY 2233 AAGGGAAGAAGGTGGGGGAGCTGCCCTCCAGCCCGAGGATACACCGAGGAAATAGTT 2292  
Db 1221 er-----CysProGlnPro-----ThrGlnProLleThr 1231  
QY 2293 TGTCTGTGCTGTGTGCAGCG-----T 2313  
Db 1231 hrArgLeuProLleThrThrGlnLysAlaArgCysTyProGlySerThrAlaProGluC 1251  
QY 2314 GTGAACCTCCCTGGGCGCTTGCCTATCCAGGCGCTCTCCCTGTCTTCTCCCTTCTT 2373  
Db 1251 yGlnProAlaThrThrProThrPro-----Thrs 1261  
QY 2374 CCAGATTATACATCTCCCTCATCTCCCTTCCCTGGGCGCCAGCGCTCCCGAGGGTTGG 2433  
Db 1261 erThrArgProLleThrThrLleThrThrSerLysProAsnCysTyProGlySerThrA 1281  
QY 2434 AAGGGCTCTGGCGCTCTTCCCTATACCATGCTGTCTTCCATAGCTTCTTCTCTCTCTCT 2493  
Db 1281 spArgArgCysProGlnLle-----ProValThr 1291  
QY 2494 TCATGAGACTGCTCCATTTCTTCTTCTCTGCAACCTGCTCCTATACAGTGAAACCTTCT 2553  
Db 1291 hrThrLys---ProArgCysTyProGlySerThr-----AspProGlnC 1305  
QY 2554 TTGGAGTGTAGTAGTACCGCTCTCTCCCGAGCCCTCAGCTGGGGCGCTGGGTGTG 2613  
Db 1305 yGlnProAlaThrTyLeuProPro---ThrThrValThrThrValArgProArgCysT 1324  
QY 2614 TCAGCGGCAATGGGGCTGTGGTTCATATGAGCGACTCTCATCTCTCTCTTGTCTCTGT 2673  
Db 1324 yr-----ProGlySerAsnAspPro-----GlyC 1332  
QY 2674 GCAGAAACCTTTGCTTCACTCCAGTGCCTCTCTAGTCCCGACCC-----TTT 2724  
Db 1332 yGlnProProGlnThrThrLysThrProLleThrThrSerLysProValCysTyLeuG 1352  
QY 2725 TCTCTCTGCTTCCCTGCAATTTCTCCAGAGAGTGTGTCTACACCT----- 2774  
Db 1352 lySerThrAspSerArgCysArgGlnAsnProSerThrThrGlnLysProLysCysSerV 1372  
QY 2775 -----CTGCTCCACTCTCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2796  
Db 1372 alGlySerThrAspProGluCysGlnProAlaThrTyLeuProProThrThrValSerP 1392  
QY 2797 CACTCACTTCTTAACCCCTGCAATCTGGCTTCCAGGCGCCAGCAATGTTCTCTCCAAG 2856  
Db 1392 roThrLleProThrThrLysProArgCysTyProGlySerThrAspThrSerCysGlnP 1412  
QY 2857 GTCGTGAGGACCTCTTCCCAAGCCCGA-----CAGTGT 2892  
Db 1412 roLysThrTySerProProThrSerArgProLleThrThrSerLysProArgCysT 1432  
QY 2893 TTGAGGCTCATCTCTCTGCTCTGTTTTCAGCCACACTGCTGAGCGCTGCTGCTT 2952  
Db 1432 yrProGlySer-----ProAspProGluCys-----GlnP 1442  
QY 2953 CTCGAACCTCTCTCTTGTCTCTGCTGCTCTCTCTGCGCCACCTTCTCTCTCTCTCTCA 3009  
Db 1442 roAlaThrTyLeuProProThrThrArgArgThrThrLleProValThrLysProArgC 1462

QY 3010 CTCCTCCAGGCT 3069  
Db 1462 yTyProGlySerSerAspProGluCysGlnProGluThrTySerProProThrThrA 1482  
QY 3070 GCCCACCCAGGCAATCAGACGCTCT-----TCCTGAGCCTCTGTGTG 3111  
Db 1482 rgProProValThrThrSerLysProAsnCysTyProGlySerThrAspSerArgCysP 1502  
QY 3112 CGTCT 3168  
Db 1502 roGlnLysProProThrThrLeuGlnProLysCysTyProGlySerSerAspProGluC 1522  
QY 3169 GCTTCAACTGTCACTGCATACAAATGATATCTCTTATGGAAAACT----- 3215  
Db 1522 yLeuAsnCysTyProGlySerProAspProArgCysProLysLleProThrThrLysL 1542  
QY 3216 -----CAGGAGGCCATGAACAAAGAGCCCTAGCATGGAGACAGGGCCAGT 3261  
Db 1542 yAlaGlyCysPheGluGlySerAspAspProArg-----C 1554  
QY 3262 GTCAGGGGACACAAAAATAGAACTTTGGGAGCAGGTATCTCTTGGTGGTGAGCCAGC 3321  
Db 1554 yGlnProAlaThrTyLeuProProSerSerArgArgProProThrThrLleAlaProLysP 1574  
QY 3322 GGCTCTGCT 3372  
Db 1574 roArgCysTyProGlySerThrAspProSerCysProGlnProThrGlnProThrThrt 1594  
QY 3373 CTGGGCGCCCTTTTCAAGAGCCAGACTTCAACCTCTCCCGCTGGGGGCGAGAGACTCT 3432  
Db 1594 hrArgThrProLleThrThrSerArgProArgCysTyProGlySerThrAspArgGluC 1614  
QY 3433 GCCCAGGGCCACTGTGTGAGTGGAGGTGGGAGAGCTCAATGACAAACTGCGGTGAC 3492  
Db 1614 yGln-----P 1616  
QY 3493 CCAGAACT---ACGAGGCTACAGCACCTTCTGTGTACTTTCGTGGCTCAACCGTCA 3549  
Db 1616 roAlaThrTySerProProThrSerArgSerGlnLleThrThrLeuLys----- 1632  
QY 3550 GGCTGCCACTGCTGAGCTGGCGGCGAGCTGGCCCACTTCTGCACCGCTTCCAGG---- 3605  
Db 1633 -----ProArgCysTyProGlySerLysAspP 1642  
QY 3606 --GCCTGTGGGAGCATTTGGGGGTGCATGGCAGCTCTGGGCTACCCACTGCCCGAGCC 3663  
Db 1642 roGluCysGlnPro-----AlaThrTyAlaPro----- 1651  
QY 3664 GCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCCA-----GTGA 3711  
Db 1652 -----ProThrThrArgThrProVal 1659  
QY 3712 CTTCTCCAGAAATGGAGACTTCTGGCTGTGAAGAGCTGCAGACTGGCTGTGGCG 3771  
Db 1659 hrThrSerLys-----ProAsnCysTyProGly----- 1668  
QY 3772 CTCGGCCAAAGACTTCAACCGCTCAAGAAGATGAGCTCCAGCAGCTGCAGCTCAC 3831  
Db 1669 -----SerThrAspLysArgCysProGlnLleValThrThrGlnLysP 1684  
QY 3832 CC---TGCACTGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCT 3888  
Db 1684 roArgCysTyProGlySer-----ProAspProGluCysGlnPro- 1697  
QY 3889 TCAACCTCTCTCCACTTTGTGAGAGCCAGCCCTGTATGCCAACACTTGTGAGCA-- 3946  
Db 1698 -AlaThrTyLeuProProThrThrSerArgThrThrLlePro---ProThrArgProAr 1716  
QY 3947 -----GGAGACAGAGCTGTGAGCTCTGGCCCTTCTCTGAGCCCGCTG 3990  
Db 1716 gCysTyProGlySerSerAspProSerCysGlnProThr-----ThrAsnLe 1732  
QY 3991 GGCGTGTGATCGATCAGCCCTGTCTCTCTCCCACTCTCTCTCTCTCTCTCTCTCTCTCT 4050



Db 131 oHis----- 132  
QY 339 CTTAAACCTAACCTAGCCCAATATATCTACCTAGCCAAACCCCTAACCTGCTT 398  
Db 132 ----- 132  
QY 399 TGCAGTCCAAAGTGTCCACTGAATCCTCAGCTTGTGCTCACTGAAATCCAGAAAG 458  
Db 133 -----ProLysProThrLysProHis-----ProHisProLysProThrLys 148  
QY 459 CATATTCCCACTGCCCACTCCCTCTTACAGACCAAC-----CTGGCCTTG 511  
Db 148 sHisHisProHisProLysProProThrLysProProLysProProSerVally 168  
QY 512 GACTCTGTGTATCTGGGATGTCAAACTCTGCAGTGCATCAGCAACCAAGCCGACTC 571  
Db 168 sProPro-----ProSerThrProLysProProTh 178  
QY 572 GTCAATGACCTCTCTCCCTTCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCT 631  
Db 178 rThrAsnProProSerThrProGlnPro----- 188  
QY 632 TGAAGTCCAACTTTTCCCGACCTTAACACCAAGAGGGGTGAACCTCCACACTGCCAC-- 689  
Db 189 -----ProThr-----HisLysProProProCysThrPr 198  
QY 690 -----CGTTTCCTGAGAGTGAGCACTAAATCTCTTCAATCTAACCACCCACC 736  
Db 198 oThrProProValAlaSerProProMetAlaThrProProThrGlnMetProProIleAl 218  
QY 737 TACACTTCCCACTCAGGAATCAGCTCCTAGAAATATACCAAACTAAGGCCCAATAGG 796  
Db 218 aThrProProIleAlaLysSerProValAlaThrProProIleAlaThrProProThrAl 238  
QY 797 CAGCCGACCTAGTGTCTAAACCTATACCTGCTTCTTCTGAGTGGTCTGCTTCTG 856  
Db 238 aThrProProIleThrIleProProValAlaThrProProIleThrProProIleAl 258  
QY 857 CGCGCCCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 916  
Db 258 laAsnProProIleMetProProIleAlaThrProProValAlaAlaProProIleT 278  
QY 917 CA---TGTGCTCTCCCACTCTGACTCCCTCAAGCTGAGCTGGAGCTGGAAGACTG 973  
Db 278 hrAsnProProIleSerLysPro-----ProVal----- 287  
QY 974 GCAGGAAGCTAGGTACAACTGGAACACGAGGCTGAGCTGAGTCCCTAGGCTGGC 1033  
Db 288 -----ThrThrProProThrThrThrProProIleAlaLysP 300  
QY 1034 CCGTCTCTCCATGTACACATATACATGTGGCACACACAGTGGCACATGCCAA 1093  
Db 300 ro-----ProIleAlaThrProPro- 306  
QY 1094 AGACTCTCTCAGCTGACACAGATCCATCTCAAGTATCTAGTAGACACTCATGCG 1153  
Db 307 -----IleSerThrProAlaAlaThrP 315  
QY 1154 TGCCAAGTCTCATCTCAACATACACATGC-----CTCTCTTCT 1195  
Db 315 roProAlaAlaThrProProIleThrThrLeuProProAlaLysProProValAlaIleS 335  
QY 1196 CTCCCTCTTG-----CCAGGAGTGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1246  
Db 335 exProIleValThrProProValThrProIleAlaGlnProProValAlaThrProProT 355  
QY 1247 CTGGTCTCCACCC-----TCACCCCAACCCAG----- 1275  
Db 355 hrAlaThrProProValAlaThrProProIleAlaThrProProThrSerLysSerProI 375  
QY 1276 -----CCCAAGTGGGGACACAGACCTGAGGGGTGCCA----- 1309  
Db 375 leSerThrProProIle-SerGluSerProValAlaThrProProThrAlaThrSerPro 394

QY 1310 -----GCTGCTTCCCGTGTGGGCCCG 1336  
Db 395 lleLysThrProProProAlaLysProProValAlaThrProProIleAlaLysSerPro 414  
QY 1337 CGCTCATGCTTCTCGTCCATCTGCCCACAGGGGACTCGTGGGGATGTTAGCTGCTG 1396  
Db 415 IleAlaThrProProThrAlaThrProPro-----Val 425  
QY 1397 TGCAGGTGCTCTGCACCTCTCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGG 1456  
Db 426 AlaThrProProIleGluLysProProValAlaThrProProThrThrProProTh 445  
QY 1457 CTGGCCCTCTCATCCAGAAACCTATGACCTCACCCGCTACCTGAGCAACCACTCCGC 1516  
Db 445 rAlaThrProProValAlaLysProProValGluThrProProIleAlaThrProProTh 465  
QY 1517 AGCTTGGTGGACCTATGTGATATCCAGCGTAGAATCTGGGAGTTGGGGAGGTGA 1576  
Db 465 rAla----- 466  
QY 1577 GGAGTTGGGAAGACAGACTCCTAACCGTGGAGGGTCTGTAATATGATGGGTGAGGAG 1636  
Db 466 ----- 466  
QY 1637 GGCTCTTGGTCCCAACCACTG-----CCCTGTCTGTCTATCTCTCCCTTCC 1687  
Db 467 -----LysProProIleSerThrProProIleSerLysProProValAlaThrPr 483  
QY 1688 CTCTTAGTGGCCCCCACTTCCCATCTCTGGGCCAGGACTAGGCATGTGGGAGGC 1747  
Db 483 oProAlaAlaThrProProIleThrThrProProPro-----ValLysProProVa 500  
QY 1748 CTGCACCGCTTGGCCCATCTGCCCTGCT----- 1781  
Db 500 lAlaThrProProLeuAlaIleProProValAlaLysProProValValThrProProTh 520  
QY 1782 -GCCAGCCAGCGCCCTCCCTCTGGGGCCG---GGGAAGTCTCTCTCTTTACAC 1837  
Db 520 rAlaThrPro-ProIleAlaThrPro-----ProIleAlaLysSerProValAlaThrP 538  
QY 1838 CGTGTGTGTGTCTCTTGGCGGGGGTGGGTGGGACAGAGGGGCCCACTCC 1897  
Db 538 ro-----ProThrAla 542  
QY 1898 ATGCTGCGT-----CCAGCTCGCTCTGCCCCAGACCT 1933  
Db 542 hrProProValAlaThrProProIleAlaLysProProValValThrProProThrTh 562  
QY 1934 GGGCCCTGCTCTGACCCAGGGCTCTCCCTTCCGCTCTCTCTCTCTCTCTCTCTCT 1993  
Db 562 hrProProThrAlaThrProProValAlaLysProProValAlaThrProProIleAla 582  
QY 1994 GGCCTCTAGGGG-----GTCAATGGGGAAGGGGACTGTAGG 2032  
Db 582 hrProProThrAlaLysProProIleSerThrProProIle-SerLysSerProValAla 601  
QY 2033 AACCCAGCAGTGTGGGAGGGGTTTAGGTGTGATGGAGTTATGCTGTAAGATTT 2092  
Db 602 ThrProProAlaAla----- 606  
QY 2093 GGGGTGTCTCAGAGGCTTTCAGAGAGCCAGGAGGAAGAGGGTGGAGGAGCC 2152  
Db 607 -----ThrProProIleThrThrProPro-----ProAla 616  
QY 2153 GAGGACCATGGGAACCGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2212  
Db 617 LysProProValAlaThrProProIle-AlaThrProProIleAlaLysPro----- 633  
QY 2213 CTCTGGAGCCAGGGAAGAAAGAGGAGGTGGGGGGAGCTGGCTCCAGCCCGAG 2272  
Db 634 -----ProValAlaThrProProTh 640

QY 2273 ATACACGGAGGAATAGTTGTTGCTGTGCTGTGCTGACGGTGTGAACCTCCCTCCCTGGGCC 2332  
Db |||||:::  
640 rAlaThrProProleAlaThrSer-----ProValAlaLysProProValAlaTh 657  
QY 2333 CTTGCTATCCAGCGCTCTCCCTGTTCTCCCTCTTCTTCCAGTATACATCTCCCT 2392  
Db |||||:::  
657 rProProle--LysThrProPro-----ProAlaLysProProValAlaLysPro 673  
QY 2393 CATCCCTTCCCTGGGCCCGCCAGCGCTCCCGAGGGTGTGAAGGCTCTGCCCTCTTC 2452  
Db |||||:::  
674 -----ProleAlaThrProProValAlaLysProProValAlaThr 687  
QY 2453 CCTATACCATGCTGTCTCCATAGCTCTCTCCCTGCTCTACTCATGAGCTGCCTCCATT 2512  
Db |||||:::  
688 Pro--ProThrAlaThrPro--ProleAlaThrProProleAlaThr--ProProVa 705  
QY 2513 TCTT-----CCTTCTCAACCCCTGCTCTATCAGCTGAACCCCTTCTTCGGAGTGTAG 2566  
Db |||||:::  
705 lValThrProProThrAlaThrProProValAlaThrProProleAlaLysProProTh 725  
QY 2567 TGAGTACCGCTCTCTCCCGCCCTCAGCTGGTGGGCTGGGTGTCTCAGCGCAATG 2626  
Db |||||:::  
725 rThrleProProThrAlaThrProProValAlaMetPro--ProleAlaThrProProT 745  
QY 2627 GGGCTCTGGTTCAATGGCCACTCTCATCTCTCTCTTGTCTTCTTGTGCGAGAAACCTTT 2686  
Db |||||:::  
745 hrAlaLysProProleAlaThrProProleAlaLysProProValAlaLysProProV 765  
QY 2687 GCTTCACTCCACTCCCTCTCTAGTTCGGACCCCTTTTCTCTCTCT----- 2732  
Db |||||:::  
765 alValThrPro--ProThrAlaThrPro--ProleAlaThrProProleAlaLys 783  
QY 2733 -----GGCTTTCCTGCCAAATTTCTCCAGGAGTGGTCTACACCTCTG-----CC 2779  
Db |||||:::  
783 rProValAlaThrProProThrAlaThrProProValAlaThrProProleAlaLysPr 803  
QY 2780 TCCACTTCTCTCCACCCACTCTCTTAACCCCTCTGCAATCTGGCTTCCAGGCCCCAG 2839  
Db |||||:::  
803 oProValAlaThrProProThrAlaThrProProThrAla----- 816  
QY 2840 CAATGGTTCTCCAGGCTCAGGCACTCTCCGCACTCTCTTCCCAAGCCGACAGTGTGGAAG 2899  
Db |||||:::  
817 -----ThrProPro----- 819  
QY 2900 CTCATTCTCTGTCTGTTTGGACCCACTCTGCTGAGCGTCTGCTCTCTCGAAC 2959  
Db |||||:::  
820 -----ValAlaLysPro-----ProValAlaTh 827  
QY 2960 TCCTCTTCTTGGTCTGCACTCTCTCTGGGCCACTTCTACCTCTCCA-----GC 3010  
Db |||||:::  
827 rPro--ProleAlaThr-----ProProThrAlaLysProProleLeuTh 842  
QY 3011 TCCTCAGGCTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3070  
Db |||||:::  
842 rProProleSerLysProProValAlaThrProProAlaAlaThrProProleThTh 862  
QY 3071 CCCACCCAGCAATCAGCAGGCTCTCTCTGAGCGCT-----TGTCGCTCTCTCTCT-- 3122  
Db |||||:::  
862 rProPro-----ProAlaLysProProValAlaThrProProleAlaThrProProIl 880  
QY 3123 -----CCTCTTTTCTACGCTCTCCATTTGGAGAGCTCACA 3160  
Db |||||:::  
880 eAlaLysProProValAlaThrProProThrAlaThrPro--ProleAlaThrSerPro 899  
QY 3161 -----CGCCACTGCTTCACTCTCACTCTCACTCTCACTCTCACTCTCACTCTCA 3196  
Db |||||:::  
900 ValAlaLysProProValAlaThrProProleLysThrProProProAla----- 916  
QY 3197 TATCCTTATTGAAAACTCAGGAGGCTGAACAAAGAGCTAGTACGAGACAGG 3256  
Db |||||:::  
917 -----LysProProValAlaLysProProleAlaThrPro 928  
QY 3257 CCAGTGTCCAGGGGACACAAAAAATAGAACTTTGGGAGCAGGTATCTCTCTGGTGTGAG 3316

Db |||||:::  
929 ProValAla-----Lys 932  
QY 3317 CAGCGGCTCTGCT 3376  
Db |||||:::  
933 ProPro--ValAlaThrProProThrAlaThrPro----- 943  
QY 3377 GGGCCCTTCTCAACGAGCCAGACTTCAACCTCTCCCGCTGGGGGAGAGACTCTGCC 3436  
Db |||||:::  
944 -----ProleAlaThrSerProleAlaThrProProValAlaThrProProThrAlaTh 962  
QY 3437 AGGCGCACTTGAAGTGTGGAGGCTCAATCAAACTGCGGCTGAGCCAG 3496  
Db |||||:::  
962 rSerProValAlaThrProProleAlaLysPro-----ProTh 975  
QY 3497 AACTAGAGGCTTACAGCCACTTCTGTGTACT-----TGCCTGGCTCAAC 3544  
Db |||||:::  
975 rThrThrProProThrAlaThrProProValAlaMetProProleAlaThrProProTh 995  
QY 3545 CGTACGCTGCCACTGTGAGCTGCGCGCAGCTGCGCCACTTCTGCACAGCTCCAG 3604  
Db |||||:::  
995 rAlaLysProProValAlaThr--ProProleAlaAsnProProValLysProProV 1015  
QY 3605 GGCCTGCTGGGAGCATGCGGCTCATGGCAGCTCTGGCTACCCAGCTGCGCCAGCG 3664  
Db |||||:::  
1015 alAlaThr--ProProleAlaLysProProThrValLeuPro--ProleAlaLysPro 1033  
QY 3665 CTGCTGGGAGTGAACCCACTTGGACTCTGCGCTGCGCCAGCTCTCTCCAGAG 3724  
Db |||||:::  
1034 ProValLysThrSerProThrAlaThr-----Pro--ProThrAlaThr--ProPro 1049  
QY 3725 ATGACGACTTCTGGCTGCTGAAGGAGCTGCGAGCTGGCTGGCTGCGGCCAGGAC 3784  
Db |||||:::  
1050 -----ValAlaLysProProValLysProProValAla-----I 1062  
QY 3785 TTCACCGCTCAAGAGAGATGCGCTCCAGCAGCTGCGCTGAGTCAACCTGCGCTGGG 3844  
Db |||||:::  
1062 leProProleThrLysProProValAlaThrProProValThrAsnProProThrAla 1081  
QY 3845 GCTCATGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3904  
Db |||||:::  
1082 -----MetProProleValThrProProp 1090  
QY 3905 CTTTGTGAGAGCCCTCTATGCCAACACTGTTGAGCCAGGAGAGAGAGTGTGAG 3964  
Db |||||:::  
1090 roileValThrPro--ProleAlaLysSerProleAlaThrProProValSerThrPro 1109  
QY 3965 CTTGCGCTTCTCTGAGCGGCTGGCGGTGTGATGATGATGATGATGATGATGATGATG 4024  
Db |||||:::  
1110 ProleAlaLysProProleAlaThrProProValAlaThrProProleAlaLysPro 1129  
QY 4025 CTTCCCAAGGCTTACCGAGCTGGGAGGAGTACAGTAGGCTCTCTCTCTCTCTCTCTCT 4084  
Db |||||:::  
1130 ProleAlaThrProProThrAlaAsnProProValAlaAsnProProleAla----- 1147  
QY 4085 TACAGGAGTCTGCTGAGGAGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 4144  
Db |||||:::  
1148 -----LysSerProleAlaLysProProleAlaThrPro----- 1159  
QY 4145 GCCTCTCTCTCTGCTTACCTTGGCCA--GTGCCCCACCCAGCCCTCAGGTGGCAG 4201  
Db |||||:::  
1160 -----ProThrAlaMetProSerleAlaThrProProle----- 1171  
QY 4202 ATCTGAGGCGGAGGTTGAGGCGCCACCCAGCAGTCTCTCTCTCTCTCTCTCTCTCT 4261  
Db |||||:::  
1172 -----GlyLysProPro----- 1175  
QY 4262 TGGCTGCCCACTCTCTCTGAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4318  
Db |||||:::  
1176 -----ValAlaThrProProMetAlaLysProProMetAlaLysSerProProleAlaThr 1193  
QY 4319 CATCCAAATCAGGAACAAACATGCTGGCAATCTTACACAAAAAGAGATGAGATTAACAG 4378  
Db |||||:::

```

Db 1194 ProProleilleLysProProVal----- 1201
QY 4379 TGcAGGgTtGGGgTCTGcATtGGAGgTCCCTATAAcCAGAGAGAAATACTGAAAGC 4438
Db 1201 ----- 1201
QY 4439 ACAGGGcAGGgACAGACAGcCAGACcCAGAGTCTCCAAAGcACAGAGTGGCAACA 4498
Db 1202 ----- 1202
QY 4499 AAACCCAGcTGAGcATcAGACcCTTG-----CCTCGAATtGTCTTCCAGTATtACGGT 4552
Db 1208 LysProValAlaThrProProValAlaThrProProValAlaLysProProValAla 1227
QY 4553 GCCTCTCTCTG-----CCCTCTTCCAGGgTATCTGTGGTtGCAGcCTGGGGAGG 4606
Db 1228 ThrSerProleilleLysProPro-----ValAlaLys 1238
QY 4607 GCACcATcCCcACcACcAGcAGcATtCTCTGAAGtTTAcATGcAGTAgcATtTTGGGGT 4666
Db 1239 ProProValThrPro----- 1244
QY 4667 GTAGGgTGcAGcCTCCcAGcGcCTCCcCCcAGcCCcACcCTcATGAcTCTAAgTG 4726
Db 1245 -----ProValAlaThrProPro----- 1250
QY 4727 TGtTGtATtAAATtTTATtTTGGAGATGtTATtTTATtTAGATGATtTTATtTGcAGAA 4786
Db 1250 ----- 1250
QY 4787 TTTCTATtCTGTATtAAcAAATAAATGCTTGCCcAGAcTtAGTCTCTTtGCCcAGC 4846
Db 1251 -----IleValLysProProIle 1256
QY 4847 CTCACcCTCTCTGtGTCTATcAGAcTCTTGcGcAcCTGGcCTCCcACTCTCTGCTtGCC 4906
Db 1257 ValThrProPro-Ile-----AlaThrProProIleAlaLysSerProIleAlaPr 1273
QY 4907 TCTGGtGGAGcTGcACAGAGcCTCTGGAGAGcGCTT----- 4946
Db 1273 O-----ProProleilleLysProProValAlaLysProProValAlaLys 1288
QY 4947 -----CCTCCcCCcACTGGGcGcAGTGGcGcACcCTcAGAcTtACCC 4987
Db 1288 rProProThrAlaThrProProValAlaThrSerProIleAlaLysPro-----Pr 1305
QY 4988 ACTGCTGTcGCCAcCACCcA-----ACcCTTGTATcCTcAGTCTCCcACAGcCTTCTGTc 5044
Db 1305 OValAlaThrProProProAlaThrProProValAlaLysProPro-----ValAl 1322
QY 5045 CACCCcAGGtTTCCCTcACCCcACTTtTGCTAAgCTCTCTCT 5085
Db 1322 aThrPro-ProThrValThrProProValAlaThrProPro 1335

RESULT 10
Q95JC9 PRELIMINARY; PRT; 676 AA.
AC Q95JC9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Basic proline-rich protein.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PAROTID GLAND;
RA Zhang Q., Szalay A.A., Kye-yune-Nyombi E., Sands J.F., Oberg K.C.,
RA Tieche J.-M., Leonora J.;
RT "Cloning and expression of a novel proline-rich protein from porcine
parotid glands.";

```

```

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY05849; AAK61383.1; -.
DR PRINTS; PR01582; KV33CHANNEL.
SQ SEQUENCE 676 AA; 62297 MW; 3008BC41EFD81FC9 CRC64;

Alignment Scores:
Pred. No.: 4,36e-18 Length: 676
Score: 373.00 Matches: 253
Percent Similarity: 27.26% Conservative: 15
Best Local Similarity: 25.74% Mismatches: 312
Query Match: 3.95% Indels: 404
DB: 6 Gaps: 41

```

US-09-931-704-3 (1-5087) x Q95JC9 (1-676)

```

QY 1156 CCAAGTCTCTAFCCTCAAAACATACATACATCCCTCTCTTTCTCTCCGCTTTGCCAGGAGTG 1215
Db 86 ProGlyProProPro-----GlyProAlaProProGly--A 98
QY 1216 TTTCCCTCTCTCCATCCCTCTGCTCCATCTGGTGTCCACCTCACCCTCCACCCACCCAG 1275
Db 98 laArgProProProProProPro-----GlyPro--ProProProGlyProAl 115
QY 1276 CCAAGGTGGGACAGACACCTGAGGGGCTGCCAGCTGTCTCCCTGTGGGCCCGGCC 1335
Db 115 aProProAlaArgProProProProProProProProProProProProProPro 135
QY 1336 GCGCTCATGCTTCTCGTCCATCCTGCCACAGGGGACTCGTGGGGATGTAGCGTGCCT 1395
Db 135 O-----AlaProProGlyAla----- 140
QY 1396 GTGCAGGTGTCTCTGGCACCTCCCTGCAGTGCAGCTCTCAATCGCACAGGGACCCAGG 1455
Db 141 -----ArgProProProProProProProProProProProProProPro 155
QY 1456 GCCTGGCCCTCCATCCAGAAACCTATGACCTCACCCCTACCTGGAGCACCACCTCCG 1515
Db 155 yPro-AlaProProGlyAlaArgProPro-ProGlyProProProProProProPro 174
QY 1516 CAGCTTGGCTGGACCTATGTAGTATCCAGCTAGGAATCTGGGAGTGGGGAGGAGTG 1575
Db 174 ----- 174
QY 1576 AGGAGTTGGGAAAGACAGACTCTAACCTGGAGGTTCTGGTAAATGATGGGTGAGGAG 1635
Db 175 -----ProGlyProAlaPro-ProGlyAlaArg-- 183
QY 1636 GGGCTTTGGCTCCACAGTCCCTCTGTGTCTATCTCTCTGCTTCTCTCTCTTAGG 1695
Db 184 -----ProProProGlyProPro----- 189
QY 1696 TGGCCCCCCTCTCCCTCCCTGGCCCGCAGCAGCCCGCTCCCTCCCTGGGGCGCG 1755
Db 190 -----ProProProProProProPro-----AlaPr 200
QY 1756 CGCCTTGGCCCATTTGCCCTCTGGCTGCAGCCCGCAGCCCGCTCCCTCCCTGGGGCGCG 1815
Db 200 oProGlyAlaArgProProProProProProProProProProProProPro 219
QY 1816 GGAAGTCTCTCTGTTTACACCGTGTGTGTGTCTCTGTGGCGGGCGGGTGGGTGG 1875
Db 220 -----AlaProProGlyAlaArgPr 226
QY 1876 GGACAGAGGGCCCTCCCTCCATCGCTCGTTCAGCTCGCTCTCCCTCTCCCTAGCTGG 1935
Db 226 oProProGlyProProProProPro-----ProProPro-- 238
QY 1936 GGCCTGTCTGTGGACCCAGGGGCTCCCTTCCGTCTGTCTCTCTCTCCCTAGCTGG 1995
Db 239 GlyProAla-----ProProGlyAlaArgProProProProProProLeuGlyPro 256
QY 1996 CCTCTAGGGGGTCTATGGGGAGGAGGACTGTAGGGAACCCAGGAGTGTGGCGAGGG 2055

```

Db 257 ProProGlyProAla-ProProGlyAlaArgProProGlyProProProGly 276  
QY 2056 GTTTAGGGTGGATGAGGTTATGCTGAAGATTGGGGTGGTCCAGAGGTGTTCCAG 2115  
Db 276 y----- 276  
QY 2116 AGAGCCAGGAGAGAGAGAGGGTTGGAGGAGCGAGCCACCATGGGAAACCGGCC 2175  
Db 277 ---ProProProGlyProAlaProProGlyAlaArgProProGly-ProProP 295  
QY 2176 CTTCTCCCGTTCCTCTCCATCCAGACCTACTCTGGAGCCAGGAAAGAAAG 2235  
Db 295 roProGlyProProPro---ProGlyProAlaPro---ProGlyAlaArgProProG 313  
QY 2236 GGAAGAGGTGGCGGGGAGCTGGCTCCAGCCCGAGGATACACCGAGGAATTAGTTGT 2295  
Db 313 lyProProProProGlyProProProProGlyPro----- 324  
QY 2296 CTTCTGCTTGTGACGGTGAACCTCCCTGGGCGCTTGCCTATCCAGGCTCTCCC 2355  
Db 325 ---AlaProProGlyAlaArgProProProGlyPro---ProProProGlyProProP 342  
QY 2356 CTTGCTTCTCCTCTCTTCCAGTTATACATCTCCCTCATCCCTTCCCTGGGCGCCAGC 2415  
Db 342 ro-----ProGlyProAlaP 347  
QY 2416 CGCTCCCGAGGTTGAAAGGCTCTGCCCTCTCTCCCTATACCATGCTCTTCCATA 2475  
Db 347 ro-----ProGlyAla----- 350  
QY 2476 GCCTTCTCTCTCTACTCATGAGACTGCTCCATTTCTTCTTCTGCAACCTGTCTCC 2535  
Db 351 ---ArgProProGlyProProPro----- 358  
QY 2536 TATCAGCTGAACCTTCTTTCGGAGTGTAGTAGTACCTCTCTCCAGCCCTCAG 2595  
Db 359 -----ProGlyProProProGlyP 366  
QY 2596 CTGGTGGCCCTGGGTGTCTCAGCGGCAATGGGGCTCTGGTTCCAAATGGGCACTCTCAT 2655  
Db 366 roAlaProProGly----- 370  
QY 2656 CTCTCTTGTTCCTTGTGCAAAACCTTTGCTTCACTCCACTGCCCTCTCTAGTCTCC 2715  
Db 371 -----AlaA 372  
QY 2716 GACCTTTTCT 2775  
Db 372 rgPro-----ProProGlyProProProProProProProProGly-----ProA 388  
QY 2776 TGCTTCCACTTCTCTCTCACCACCTCACTTCTTAACCCCTGCAATCTGGCTTCCAGGCC 2835  
Db 388 laProProGlyAlaArgProProProGlyProProProProGlyProProProGlyP 408  
QY 2836 CCAGCAATGGTTCCTCAAGTCTGTCAGGACCTCTCTTGCACCCGACAGTGTTTG 2895  
Db 408 roAlaProProGlyAlaArgProProProProProProProProAlaaspGluProGIng 428  
QY 2896 AAGGCTCATCTCTCTGTCTGTCTTTGTCAGCCACATGCTGAGCGGTCTGCTCTCTCT- 2954  
Db 428 lnglyPro-----AlaProSerG 434  
QY 2955 -----CGAATCT 2997  
Db 434 lyAspLysProLysLysProProProProProAlaGlyProProProGlyProProS 454  
QY 2998 CTACTCTCCAGCTCTCTCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3057  
Db 454 exProGlyProAlaProProProGlyAlaArg-----ProProProGlyProP 469  
QY 3058 CTCCCAAGTTTGGCCACCCAGCCCAATCAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3117  
Db 469 roProProGlyProProGlyProAlaProProGlyAlaArgProProProGlyP 489

QY 3118 CT 3167  
Db 489 roProProProGlyProProProProGlyProAlaProProGlyAlaArgProProPro- 508  
QY 3168 TGCTTCACTGTCACTGTCATCAATATGATATCTTATTGGAAAACTCAGGAGGCCAT 3227  
Db 509 -----GlyProPro 512  
QY 3228 GAACAAAGAACCTTAGCATGGAGACGGGCGAGTGTGAGGGGACACAAAAAATAGAACT 3287  
Db 512 roProGlyProProProProProGlyProAlaProPro----- 523  
QY 3288 TTGGAGCAGGATATCTCTTGTGGTGGAGCCAGCGGTCTGCCCTCTCTCTCTCTCTCTCTCA 3347  
Db 524 --GlyAlaArgProProProGlyProProProGly-ProProProProGlyProAla 542  
QY 3348 CCCTCA 3407  
Db 543 Pro---Pro-GlyAlaArgProProProGlyProPro-----ProPr 555  
QY 3408 CTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGCGAA 3467  
Db 555 oGlyProProPro-----ProGlyProAlaP 564  
QY 3468 GCCTCAATGACAACTGGGCTGACCCAGAACTACGAGGCTACAGCCACTCTCTGTGT 3527  
Db 564 oProGlyAlaArgProProProGlyProProProProGlyProProProGly----- 582  
QY 3528 ACTTGGCTGGCTCAACGGTCAGGCTGCCACTGTGAGCTGGCGGACGCTGGCCCACT 3587  
Db 583 -----ProAlaProProGly-----AlaArgProProProGlyPro- 594  
QY 3588 TCTGCACCGCTCCAGGGCTGTGGGCGCACTTGGGGCGCTCATGCGAGCTCTGGGT 3647  
Db 595 -----ProProProGlyProPro----- 600  
QY 3648 ACCCACTGCCAGCCGCTGCTGGAGCTGAACCCACTTGGACTCTCTGGCTCTGCCACA 3707  
Db 601 -ProProGlyProAlaPro-ProGlyAlaArgPro-----ProProGlyPro--ProPr 617  
QY 3708 GTGACTTCTCTCAGAGATGAGACCTTCTGGCTCTGAAGGAGCTGCGAGCTGTGGT 3767  
Db 617 oProGlyProProProProGly----- 624  
QY 3768 GCGCTCGGCCAAGACTTCAACGGCTCAAGAGAGAGATGAGCGCTCCAGCAGCTGCAG 3827  
Db 625 -----ProAla-----ProProGlyAlaArgProProProGly 635  
QY 3828 TCACCTGTCACCTGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCT 3887  
Db 635 yProProProProGly-----ProProPr 643  
QY 3888 TTCAACCT 3947  
Db 643 oProGlyProAlaProProGlyAlaArgPro-ProProGlyProProProProProG 663  
QY 3948 GAGACAGAAAGCTGTGAGCTCTGGCCCTTCTCTGACCGGCTGTGGCGGTGTGATGATCA 4007  
Db 663 ly----- 663  
QY 4008 GCCCTGTCT 4042  
Db 664 --ProSerProProArgProProProGlyProPro 674

## RESULT 11

Q9NYW9 PRELIMINARY; PRT; 1548 AA.  
ID Q9NYW9  
AC Q9NYW9;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Somatostatin receptor-interacting protein splice variant b.





[illegible]

















Db 445 ProPro----- 446  
Qy 2928 CCACACTGCTGAGCGCTGCTGCTTCTCGAAGTCTCTTCTTCTGCTGCTGCTGCTCTCTCT 2987  
Db 447 -----ArgProAlaProProProLeuProHisArgGluGlyPheLeu 460  
Qy 2988 GGGCCACCTCTCT-----ACCTCTCCAGCTCTCT 3014  
Db 461 GlyProAlaAspArgPheSerValGlyThrGlnAspSerHisThrProProThrPro 480  
Qy 3015 CCAGGCT 3071  
Db 481 ProThrProThrThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 496  
Qy 3072 CCACCCAGCAATCAGCAGCT 3131  
Db 497 -----SerSerSerAlaGlyPro-----ValSerPheProProProPro-Ty 510  
Qy 3132 TCTAGCGCTCTCCATTGGAGAGCTCACCA----- 3160  
Db 510 rLeuAlaArgSerIleAspProLeu-ProArgProProSerProAlaGlnAsnProGlnA 530  
Qy 3161 --CGCCAGCTGTTCACTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3206  
Db 530 spProProLeuValProLeuThrLeuAlaLeuProProAlaProProSerSerSerSerHisG 550  
Qy 3207 GGAATAAATCTCAGG-----AGGCCATGACAAAGAGCCCTAGCATGGAG----- 3250  
Db 550 lnAsnThrSerGlySerPheArgArgProGluSerProArgProArgValSerPheProL 570  
Qy 3251 -----ACAGGCCAGTGTCA----- 3265  
Db 570 ysThrProGluValGlyProGlyProProProGlyProLeuSerLysAlaProGlnProV 590  
Qy 3266 -----GGGACACAAATAAGAACTTTGGAGCAGGTATCTCTCTGTTGGT 3311  
Db 590 alProProGlyValGlyGluLeuProAlaArg-----GlyProArg--LeuPheAspP 607  
Qy 3312 GTGAGCAGCGCTGCTGCT 3340  
Db 607 heProProThrProLeuGluAspGlnPheGluProAlaGluPheLysIleLeuProA 627  
Qy 3340 ----- 3340  
Db 627 spGlyLeuAlaAsnIleMetLysMetLeuAspGluSerIleArgLysGluGluGlnG 647  
Qy 3341 -----CCCATCACCCT 3368  
Db 647 lnGlnHisGluAlaGlyValAlaProGlnProProLeu-LysGluProPheAlaSerLeu 666  
Qy 3369 ACTACCTGGCCCT 3428  
Db 667 GlnSerProPheProThrAspThrAlaProThrThrThrAlaProAla----- 682  
Qy 3429 CTCTGCCAGGCCACTGTTGACTTGGAGGTGGCGAAGCTCAATCACAACCTGCGCGC 3488  
Db 683 -----ValAlaValThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 698  
Qy 3489 TGACCCAGAACTACA-----GGCCTACAGCCACTCTCTGTTTACTTGGCTG 3536  
Db 698 rGlnGluGluGluLysLysProProProAlaLeuProProProProProProLeuAlaLysPh 718  
Qy 3537 GCCTCAACCGTCAGCTGCCACTGTGAGCTGCGCGCGAGCTGCGCCACTCTCTGCA 3596  
Db 718 eProProProSer-----GlnProGlnProProPro-----ProPr 730  
Qy 3597 GCCTCAGGCGCTGCTGGC-----AGCATTCGCGCGCTCATGGCAGCTCTGCGCT 3647  
Db 730 oProProSerProAla-SerLeuLeuLysSerLeuAlaSerValLeuGluGlyGlnLysT 750  
Qy 3648 AC-----CCACTGCGCCAGCGCTGCTGCGGACTG 3677  
Db 3678 AACCCACTTGACCTCTGGCCCTGCCACAGTACTTCTCTCCAGAGATGACGACTTCT 3737  
Db 770 In-----TySerProGlyPro-ProSer----- 777  
Qy 3738 GGCTGCTGAAGAGCTGCGACCTTGGTGTGGCTCTGGCCCAAGGACTTCAAC----- 3791  
Db 778 -----GlyAlaThrAlaLeuProProThrSerAlaAlaPro 789  
Qy 3792 -----GGCTCAAGAGAGATGCGACCTCCAGCAGCTGCGAGTCACTGACCTGG 3842  
Db 790 SerAlaGlnGlySer-ProGlnProSerAlaSerSerSerSerGlnPheSerThrSerG 809  
Qy 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3902  
Db 809 yGlyProThrAlaArgGluArgAla-----GlyGluGluProValProGly-P 826  
Qy 3903 CACTTTGTGAGAGCCCTGTATGCCAACACCTGTTGAGCCAGGA----- 3949  
Db 826 roMetThrProThrGlnProProProProLeuSerLeuProAlaArgSerGluSerG 846  
Qy 3950 -----GACAGAAGCTGTGAGCCCTCTGCGCTTCTCTGAGCCGCTGG 3992  
Db 846 luValLeuGluGluIleSerArgAlaCysGluThrLeuValGluArgValGlyArgSerA 866  
Qy 3993 CGTGTGATCGATCAGCCCTGCT 4052  
Db 866 laThrAspProAlaAspProValAspThrAlaGluPro----- 878  
Qy 4053 GAGGTACAGTAGGCCCTGCT 4112  
Db 878 ----- 878  
Qy 4113 AAGTGGTTGAGTGTGCTGAGCGGCTCATGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4172  
Db 879 -----AlaAspSerGlyThrGluArgLeuLeu-ProProAlaGlnAla----- 892  
Qy 4173 CAGTCCCCCAGCCAGCCCTCAGTGGGCACATCTGGAGGGCAGGGGTG----- 4220  
Db 893 -----LysGluGluAlaGlyValAlaAlaValSerGlySerCysLysArgArg 909  
Qy 4221 -----AGGGCCACACACACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4253  
Db 910 GlnLysGluHisGlnLysGluHisArgArgHisArgAlaCysLysAspSerValGly 929  
Qy 4254 ---AAGCCCTTGGCTGCCCT 4310  
Db 930 ArgArgPro----- 932  
Qy 4311 ACTCTATACATCCAATTCTCAGGAAACAAACATGTTGGTGAATTTCTACACAAAGAGATGAG 4370  
Db 932 ----- 932  
Qy 4371 ATTAACAGTGCAGGTTGGGTGTGCTGCAATTTGAGGTGCTTAAACACAGAGAGAAATA 4430  
Db 933 ---ArgGluGlyArgAlaLysAlaLysValProLysGluLysSerArgArgVal 951  
Qy 4431 CTG-----AAAGCAGGGGCGGAGGACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 4466  
Db 952 LeuGlyAsnLeuAspLeuGlnSerGluGluIleGlnGlyArgGluLysSerArgProAsp 971  
Qy 4467 --CCAGGAGTCTCCAAAGCAG----- 4487  
Db 972 LeuGlyGlyAlaSerLysAlaLysProProThrAlaProAlaProProSerAlaProAla 991  
Qy 4488 ---AGTGGCAACAAACCCAGGCTGACATCAGGACCT 4523  
Db 992 ProSerAlaGlnProThrProProSerAlaSerValPro 1004  
RESULT 15  
Q9WU13 PRELIMINARY; PRT; 2158 AA.  
ID Q9WU13

AC Q9WU13;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Synaptic SAPAP-interacting protein Synammon.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99419021; PubMed=10488079;  
 RA Yao I., Hata Y., Hixao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;  
 RT "Synammon, a novel neuronal protein interacting with synapse-associated  
 RT protein 90/postsynaptic density-95-associated protein.",  
 RL J. Biol. Chem. 274:27463-27466(1999).  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 DR EMBL; AF102855; AAD04569.2; -;  
 DR HSSP; P00519; IABL.  
 DR InterPro; IPR002110; ANK.  
 DR InterPro; IPR001478; PDZ.  
 DR InterPro; IPR002965; P-rich\_extensin.  
 DR InterPro; IPR001660; SAM.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00023; ank; 7.  
 DR Pfam; PF00595; PDZ; 1.  
 DR Pfam; PF00536; SAM; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR01217; PRICHEXTENSIN.  
 DR PRODOM; PD00066; SH3; 1.  
 DR SMART; SM00248; ANK; 3.  
 DR SMART; SM00228; PDZ; 1.  
 DR SMART; SM00454; SAM; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50088; ANK\_REPEAT; 3.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS50106; PDZ; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW ANK repeat; Repeat; SH3 domain.  
 SQ SEQUENCE 2158 AA; 225519 MW; 2AB6D5385F1A4407 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.29e-16 Length: 2158  
 Score: 354.00 Matches: 334  
 Percent Similarity: 29.90% Conservative: 102  
 Best Local Similarity: 22.91% Mismatches: 483  
 Query Match: 3.75% Indels: 541  
 DB: 11 Gaps: 69  
 US-09-931-704-3 (1-5087) x Q9WU13 (1-2158)  
 QY 1084 CACATGCCAAGACTCTC-----TCAGCTGACACAGATCCATTCTCAA 1128  
 Db 849 HisArgProLysGlyPhePheAlaThrGluSerSerPheAspProHisArgSerGln 868  
 QY 1129 GPATCTACTGATAGACACATCGCGTCCCA----- 1158  
 Db 869 ProSerTyrAspArgProSerPheLeuProGlyProGlyLeuMetLeuArgGlnLys 888  
 QY 1159 -----AGTCTCATCTCTCAAAATACACA 1182  
 Db 889 SerIleGlyAlaAlaGluAspArgProTyrLeuAlaProAlaMetLysPheSer 908  
 QY 1183 TGCCCTCTTCTCTCCCTCTGTCAGGAGTGTTCCTCTCTCCATCC-----C 1233  
 Db 909 -ArgSerLeuSerValProGlySerGluAsp-IleProProProProThrThrSerProp 928  
 QY 1234 CTCTGCTCCCTCTGCTGCTCCACCTCACCCTCCACCCACCCAGGCGGAGGAGAC 1293  
 Db 928 roGluProProTyr-SerThrProProAlaProSer-Ser-GlyArgLeuThrProSe 947  
 QY 1294 ACCTGAGGGCTGCCAGCTGCTTCCCGGTGGCGCGCGCTCATGCTTCGTC 1353  
 Db 1237 LysGluGlyGlyTrpGlnAsnGluAlaArgArgSerThrLeuPheLeuSerThrAsp 1256  
 Db 947 rProArgGlyGlyPro---PheAsnProSerSerGlyGlyPro-LeuProAlaSerSerP 966  
 QY 1354 CATCC-----TGCCACAGGGGACTCTGGGGGATTTAGCTGCT--- 1395  
 Db 966 roSerSerPheAspGlyProSerProAspThrArgGlyGly-GlyArgGlyLysSer 985  
 QY 1396 --GTGCACGGTGTCTGGCACCTCCCTGC-----AGTGCCAGCTCTCAATCGCAC 1443  
 Db 986 LeuTyrHisSerAlaAlaLeuProProAlaHisHisHisProHisHisHisHis 1005  
 QY 1444 AGGGGACCCAGGGCTGGCCCTCCATCCAGAAAACCTATGACCTCACCCG----- 1494  
 Db 1006 HisAlaProProProGlnProHisHisHisHisAlaHisProHisProGluMet 1025  
 QY 1495 -----CTACCTGGAGCACCACTCCGAGCTTGG-CTGGGACCTATGTGAGTA 1541  
 Db 1026 GluThrGlyGlySerProAspAspProProProArgLeuAlaLeuGlyProGln----- 1043  
 QY 1542 TCACGCGTAGGAATCTGGGAGTGGGAGGAGTGGGAGAGTGGGAGAAACAGACTCTCAAC 1601  
 Db 1044 ProSerLeuArgGlyTrp----- 1049  
 QY 1602 CGTGGAGGGTTCTGTAATATGATGGGTGAGAGGGGCTCTTTGGCTCCACAGTCCCC 1661  
 Db 1050 ArgGlyGly-----GlyProSerPro 1056  
 QY 1662 CTGCTGCTATCTCTCTGCTTCCCTCTTAGTGGGCGCCCTCTTCCCATCCCTGG 1721  
 Db 1057 ThrSerGly-----AlaProSerProSerHisSer 1067  
 QY 1722 CCCAGGACTAGGCATGTGGGAGGCTCGCACCCGCTCT-----TGGCCCATTTGCC 1772  
 Db 1068 SerSerGlyGlySerSerGlyProThrGlnAlaProAlaLeuArgTyrPheGlnLeuPro 1087  
 QY 1773 CCAGTGGCTGCCAGCCAGCGCCGCTCCCTCGGGGGCGGGAGTCTCTCTCTGT 1832  
 Db 1088 ProArgAlaAlaSerAlaAlaMetTyrValProAlaArgSerGlyArgGlyArgLysGly 1107  
 QY 1833 TACACCGTGTGTGTCTCTTGGCGGGGGTGGTGGGAGCA----- 1880  
 Db 1108 ProLeuValLysGlnThrLysValGluGlyGluProGlnLysGlySerIleProSerAla 1127  
 QY 1881 GAGGGGCGCCACCTCCCATGCTCGTTCAGCTCGCTCTGCCCCCAGACCTGGGGCCC 1940  
 Db 1128 SerSerProThrThrProAlaLeuProArgSerGlu-----ProProAlaGlyPro 1145  
 QY 1941 TGCTGCTCTGACCCAGGGGCTCTCTCGCTCTCTCCATCTCCATCTAGCTGGGCTCC 2000  
 Db 1146 SerGluLysAsnSerIleProIleProIleIleLysAlaProSerThrSerSer 1165  
 QY 2001 TAGGGGGTCTATGGGGAGGGGACTGTAGGGA-----ACC 2036  
 Db 1166 -----SerGlyArgSerSerGlnGlySerSerThrGluAlaGluProThrThr 1181  
 QY 2037 CAGCAGTAGTGGCAGGGGGTTAGGGTGTGAGGTGAGAGGTATAGCTGAAGGATTTGGG 2096  
 Db 1182 GlnProAspGlyAlaGlyGlyGly----- 1190  
 QY 2097 GTGCTCCAGAGGTTCAGAGAGCCAGAGAGAGAGAGAGAGAGGTGGAGAGCGGAGG 2156  
 Db 1191 -----SerSerPro-----SerProAla 1196  
 QY 2157 CACATGGGNAACCGGCGCTCTTCCGCTTCTCTCTCCATCCAGACCTCTACTCT 2216  
 Db 1197 ProAlaThrSerProValProProSerProSerProValProThrProAlaSerProSer 1216  
 QY 2217 GGA----- 2219  
 Db 1217 GlyProAlaThrLeuAspPheThrSerGlnPheGlyAlaAlaLeuValGlyAlaAlaArg 1236  
 QY 2220 -----GCCGGGAAGAGAAA----- 2234  
 Db 1237 LysGluGlyGlyTrpGlnAsnGluAlaArgArgSerThrLeuPheLeuSerThrAsp 1256



```
Db 1926 lntHrSerLeuSerLysProSerSer-----SerIlePheGlnAsnTrp 1942
QY 3698 CCTGCCACAGTACTCTCCAGAGATGACGACTTCTGGCTCCTGAAGAGCTGCAG 3757
Db 1942 rOlysProPro---LeuProPro-----LeuProThrGlySerGlyValSer 1957
QY 3758 ACCTGGCTGTGGCTCGCGCCAGGACTTCAACGGCTCAAGAGAGATGCAGCTCCA 3817
Db 1957 erSerThrAlaAlaAlaProGlyAlaThrSerProSer-----AlaSerSerAlaSerA 1975
QY 3818 GCAGCTGCAGTCACCTGCACCTGGGGCTCATGCTTCTGACTTCTGACTTCTCCTCT 3877
Db 1975 laSerThrArgHis---LeuGlnGlyValGluPheGluMetArgProProLeuLeuArgA 1994
QY 3878 TCGCTCCC---CCTTCA-----AACCCTGCTCCCACTT 3907
Db 1994 rgAlaProSerProSerLeuLeuProAlaSerAspHisLysValSerProAlaProArg- 2013
QY 3908 TGTGAGAGCCAGCCCTGTATGCCAACA---CCTGTGTAGCCAGGAGACAGAGCTGTGAG 3964
Db 2014 ProSerSerLeuProIleLeuProSerGlyProIleTyrProGlyLeuPheAspIleArg 2033
QY 3965 CCTCTGCCCTTTCCTCGACGGCTGGCGGTGTGATGGATCAGCCCTGTCTCCTCCCA 4024
Db 2034 SerSerProThrGlyGlyAlaGlySerThrAspProPheAlaProValPheValPro 2053
QY 4025 CCTCCCAAGGTCTTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTCTCTCTGTTTC 4084
Db 2054 ProHisProGlyIle-----SerGlyGlyLeuGlyAlaLeuSer----- 2067
QY 4085 TACAGGAAGTCATGCTCGAGGAGGTGTGAAGTGTTCAGTTGTGTGCAGAGCGCTCATG 4144
Db 2068 -----GlyAlaSerArgSerLeuSerProThrArgLeu----- 2078
QY 4145 GCCTCTGTCTTCTGTCTTACCACTTGGCCAGTGCACCCAGCCCTCAGGTGGCACATC 4204
Db 2079 -----LeuSerLeuProProAspLysProPheGlyAlaLysProLeu----- 2092
QY 4205 TGGAGGGCAGGGGTGTAGGGGCCACCACACATGCTTCTGGGGTGAA----- 4255
Db 2093 -----GlyPheTrpThrLysPheAspVal 2100
QY 4256 GCCCTTGGCTGCCCACTCTCCTTGGATGGGTGTCTCCTTATCCCAAAATCACTCT 4315
Db 2101 AlaAspTrpLeu-----GluTrpLeuGlyLeuSerGluHisArgAlaGlnPheLeu 2117
QY 4316 ATACATCCAATTCCAGGAACACAAACATGGTGGCAATT 4351
Db 2118 AspHisGluIleAspGlySerHisLeuProAlaLeu 2129
```

Search completed: January 27, 2003, 16:20:13  
Job time : 530.364 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:43:12 ; Search time 214.014 Seconds  
(without alignments)  
6334.602 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 9432

Sequence: 1 aactcgagtgaggcctggc.....ccttgtaagtcttctca 5087

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q/cgn2\_1/USP001/US09931704/runat 27012003 154125 3585/app\_query.fasta\_1.7189  
-DB=A Geneseq 101002 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRF=US09931704.cgn 1 1 222 @runat 27012003 154125 3585 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A Geneseq 101002:.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1446	15.3	321	22	ABBI1896 Human cardiostroph
2	1446	15.3	321	22	AAW79399 Human protein SEQ
3	1112	11.8	260	22	AAW78415 Human protein SEQ
4	910.5	9.7	253	21	AAW25831 Human protein sequ
5	896	9.5	215	21	AAW19586 Human interleukin-
6	896	9.5	225	19	AAW29715 Human neurotrophic
7	896	9.5	225	19	AAW56141 Amino acid sequenc
8	896	9.5	225	20	AAW94466 Human cardiostroph
9	896	9.5	225	21	AAW87813 Human NNT-1 protei
10	896	9.5	225	22	AAW63543 Human novel neurot
11	896	9.5	225	23	AAW78176 Peptide #7823 enco
12	885	9.4	164	22	ABBA0317 Protein #6715 enco
13	885	9.4	164	22	ABBA24716 Human brain expres
14	885	9.4	164	22	AAW61118 Human bone marrow
15	885	9.4	164	22	AAW73827 Peptide #6549 enco
16	885	9.4	164	22	AAW20115 Peptide #8049 enco
17	885	9.4	164	22	AAW34012 Human peptide enco
18	885	9.4	164	23	ABGA3716 Human cardiostroph
19	884	9.4	223	22	AAE00828 Mouse interleukin-
20	863	9.1	215	21	AAW19587 Amino acid sequenc
21	863	9.1	225	19	AAW29716 Murine NNT-1 prote
22	863	9.1	225	19	AAW56142 Mouse novel neurot
23	863	9.1	225	21	AAW87814 Human 5' EST secre
24	863	9.1	225	23	AAW78177 Mycobacterium tube
25	496	5.3	94	20	AAW12768 Drosophila melanog
26	396.5	4.2	763	18	AAW31852 Human apolipoprote
27	382.5	4.1	1047	22	ABW70874 Novel human diagno
28	378.5	4.0	2284	22	ABW71434 Novel human diagno
29	376.5	4.0	1008	22	ABW11527 Amino acid sequenc
30	374	4.0	1013	22	ABW08112 Drosophila melanog
31	366	3.9	4561	22	ABG30203 Human N-methyl-D-a
32	366	3.9	9222	22	ABG21064 Collagen type III-
33	364	3.9	2639	22	ABG15016 Collagen type III-
34	351.5	3.7	2087	22	ABG1516 Amino acid sequenc
35	341.5	3.6	2703	22	ABG63299 Drosophila melanog
36	339.5	3.6	1212	20	AAW87503 Human N-methyl-D-a
c 37	334	3.6	1078	16	AAW1704 Collagen alpha 1 (
c 38	334	3.5	2091	21	AAW12000 Collagen type III-
c 39	330.5	3.5	1078	21	AAW96125 Collagen type III-
c 40	330.5	3.5	1078	23	ABW0736 Amino acid sequenc
c 41	330.5	3.5	1078	23	ABW09628 Human collagen alp
c 42	330.5	3.5	1078	23	AAE16478 Human ORFX ORF995
c 43	330.5	3.5	2971	21	AAW41231 Novel human diagno
c 44	329.5	3.5	1469	22	ABG15191 Collagen type III-
c 45	327.5	3.5	1466	22	ABW50291

# ALIGNMENTS

## RESULT 1

ABBI1896  
ID ABBI1896 standard; peptide; 321 AA.

XX ABBI1896;

AC ABBI1896;

XX 11-JAN-2002 (first entry)

DE Human cardiostrophin-like cytokine homologue, SEQ ID NO:2366.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;

cell culture; drug screening; gene therapy; antiinflammatory; antiaschmatic; antiarthritic; haemostatic; antiarteriosclerotic; cystostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial; antifungal; vulnery; antiulcer.

**Homo sapiens.**

WO200157188-A2.

09-AUG-2001.

05-FEB-2001: 2001WO-US03800.

03-FEB-2000: 2000IS-0496914

27-APR-2000; 2000US-0560875.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-457740/49.

11-11-11

treating or ameliorating a medical condition in a mammalian subject

[illegible][illegible]

sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoietic regulatory activity; tissue growth activity;

Sequence 321 AA;

Alignment, Scores:

Pred. No.:

**Score:**  
**Percent**

recessive character:

1.2e-96

1446.00  
34 498

**O.F.C.**

Length:

### Matches

conserv.

321

319

2

```
Best Local Similarity: 34.26%      Mismatches: 0
Query Match:          15.33%      Indels: 610
DB:                   22          Gaps: 1
```

US-09-931-704-3 (1-5087) x ABB11896 (1-321)

[illegible]



PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejрман T, Goodrich R;  
XX  
DR WPI; 2001-476283/51.  
DR N-PSDB; AAK52532.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
PS Claim 20; Page 237; 6221pp; English.  
XX  
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AM78323-AM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
SQ Sequence 321 AA;

Alignment Scores:  
Pred. No.: 1,2e-96 Length: 321  
Score: 1446.00 Matches: 319  
Percent Similarity: 34.48% Conservative: 2  
Best Local Similarity: 34.26% Mismatches: 0  
Query Match: 15.33% Indels: 610  
DB: 22 Gaps: 1

US-09-931-704-3 (1-5087) x AAM79399 (1-321)

QY 1064 TTGGCACACACAGTGGGCACATGCCAAGACTCTCTCAGCTGACACAGATCCATT 1123  
Db 1 ValAlaHisThrGlnTrpHisThrCysGlnArgLeuSerGlnLeuThrHisArgSerIle 20  
QY 1124 CTCAGTATCTACTGATAGACACTCATGCTGCGCAAGTCTCTCTCTCAACATACACAT 1183  
Db 21 LeuLysTyrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40  
QY 1184 GCCTCTCTTCTCTCCGCTTTCAGGAGTGTTCCTCCCTCTCCATCCCTCTGCTCC 1243  
Db 41 AlaSerLeuSerLeuProSerCysGlnGluCysPheProSerSerIleProSerAlaSer 60  
QY 1244 CATCTGCTCTCCACCTCACCCCCACCCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1303  
Db 61 HisMetValSerHisProHisProProProSerProArgTrpGlnTrpGlnTrpGluGly 80  
QY 1304 CTGCCAGCTGCTTCCCGGTGGGCGCGCGCTCATGCTTCTCGTCCATCCCTGCC 1363  
Db 81 LeuProAlaAlaSerProCysGlyProGlyProArgSerCysPheSerSerIleLeuPro 100  
QY 1364 ACAGGGGACTCTGGGGGATGTTAGCGTGTGTGCAACGCTGTGTGGCACCTCCCTGCA 1423  
Db 101 ThrGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAla 120  
QY 1424 GTGCCAGCTCTCAATCGCACAGGGGCCCCAGGCGCTCCATCAGCTGAAACCTAT 1483

Db 121 ValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyr 140  
QY 1484 GACCTCACCCGCTACTGGAGCACCACTCCCGACGCTGGCTGGGACCTATGTGAGTATC 1543  
Db 141 AspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThrTyr 157  
QY 1544 CAGCGTAGGAATCTGGGAGTTGGGAGGAGTGGGAGTTGGGGAAGACAGTCTTAACCG 1603  
Db 157 157  
QY 1604 TGGAGGGTTCTGGTAAATGATGGGTGAGAGGGGCTCTTTGGCTCCCAACCTCCCT 1663  
Db 157 157  
QY 1664 GTCTGGTCTATCTCTGCTCCCTTCTCTTAGGTGGGCCCCCCTTCCCATCCCTGGCC 1723  
Db 157 157  
QY 1724 CCAGGACTAGGCATGTGGGCAGGCCTCGACCCGCTTGGCCCATGTGCCCACTGGCTGC 1783  
Db 157 157  
QY 1784 CAGCCAGCCCGCCCTCTCCCTCTGGGGCCGGGAAAGTCTCTGTGTTTACACCGTGT 1843  
Db 157 157  
QY 1844 GTGGTGTCTCTTGGCGGGCGGGTGGGTGGGACAGAGGGGCCCTCCCATGCCT 1903  
Db 157 157  
QY 1904 GCGTTCAGCTCGCTCTGCCCCCAGACGTGGGGCCCTGCTCTGTGACCCAGGGGCT 1963  
Db 157 157  
QY 1964 CCCTTCGCTGCTCTCTCCCATCTAGTGGGCTCTTAGGGGGTCTATGGGGGAAGGG 2023  
Db 157 157  
QY 2024 ACTGTAGGGAACCCAGGAGTAGTGGCAGGGGTTTAGGGTGTGGATGGAGGTATGCTG 2083  
Db 157 157  
QY 2084 TAAGGATTTGGGGTGGTCCAGAGGTGTTTACAGAGAGCCAGGAGAGAGAGAGGGT 2143  
Db 157 157  
QY 2144 GGAGAGCCGAGGAGCACCATGGGAAACCGGCCCTCTTCCCGTGTCTCTTCCACATCC 2203  
Db 157 157  
QY 2204 CAGACCCTACTCTGAGCCAGGAAAGAAAGGAAAGAGTGGCGGGGAGCTGCTCC 2263  
Db 157 157  
QY 2264 AGCCCGAGATACACAGGAAATTAAGTTGTCTGTGCTGTGTCAGCGTGTGAACCTCC 2323  
Db 157 157  
QY 2324 CCCTGGGCGCTTGGCTATCCAGGCTCTCCCTTGTCTTCTCCCTTCTTCCAGTTATA 2383  
Db 157 157  
QY 2384 CATCTCCCTCATCTTTCCTGGGCCCCAGCGCTCCCGAGGGTGGAAAGGGCTCT 2443  
Db 157 157  
QY 2444 GCCCTCTTCCCTATACCATGCTGTCTTCCATAGCCTTCTCTGTCTACTCATGAGACT 2503  
Db 157 157  
QY 2504 GCCTCCATTTCTCTCTTCTGCAACCCGCTCTCTATCAGCTGAAACCTCTTCTTCGGAGTGT 2563



Db 157 ----- 157  
Qy 2564 TAGTGAGTACCGTCTCTCTCCAGCCCTCAGCTGGTGGCTGGTGTGTGAGCGCAA 2623  
Db 157 ----- 157  
Qy 2624 ATGGGGCTCTGTTTCCAATGGGCCACTCTCATCTCTCTCTGTTTCTGTGGAGAAACC 2683  
Db 157 ----- 157  
Qy 2684 TTTGCTTCACTCCACTGCGCTCTCTAGTTCCGACCCCTTTTCTCTCTGCTTTCCCTG 2743  
Db 157 ----- 157  
Qy 2744 CCAATTTTCTCAAGGAGTGTCTACACCTCTGCTCTCACTCTCTCTCCACCCACTCAC 2803  
Db 157 ----- 157  
Qy 2804 TTCTTAACCCCTGCAATCTGGCTTCCAGGCCCCAGCAATGTTCTCTCAAGGTGCTCA 2863  
Db 157 ----- 157  
Qy 2864 GGCACCTCTCTGCAAGCCGACAGTGTTTGAAGGCTCATTTCTCTGCTCTCTGTTT 2923  
Db 157 ----- 157  
Qy 2924 GCAGCCACACTGCTGAGGCTGCTGCTTCTCGAACTCTCTTCTGCTGCTCTGCACTC 2983  
Db 157 ----- 157  
Qy 2984 TCCTGGGCACTTTTACCTCTCCAGTCTCTCCAGGCTCTCTCTCTCTCTCTCTGCTG 3043  
Db 157 ----- 157  
Qy 3044 CCACAGGGGCACTCTCCCAAGTTTGGCCACCAGCCCAATCAGCAGCTCTCTCTGAGC 3103  
Db 157 ----- 157  
Qy 3104 GTCTTGTGGTCT 3163  
Db 157 ----- 157  
Qy 3164 CCACTGTTCACTGTCTACCTGCATACAAATGATATCTCTTATTTGAAATACTCAGGAGG 3223  
Db 157 ----- 157  
Qy 3224 CCATGAACAAAGCCTAGCTAGTGGAGACAGGGCCAGTCTCAGGGGACACAAAAATAGA 3283  
Db 157 ----- 157  
Qy 3284 AACTTTGGAGCAGGTATCTCTTGTGTGTGAGCCAGCGCTCTGCTCTCTCTCTCTCTCT 3343  
Db 157 ----- 157  
Qy 3344 ATCACCCTCTCTTTTCACTGTAACCTACCTGGGGCCCCCTTTTCAACAGCAGCACTTC 3403  
Db 158 -----LeuAsnTyrLeuGlyProProPheAsnGluProAspPhe 170  
Qy 3404 AACCTCCCGCTGGGGCAGAGCTCTGCCAGGCGCCACTGTTCAGCTTGGAGGTGTGG 3463  
Db 171 AsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeuGluValTrp 190  
Qy 3464 CGAAGCTCAATGACAACTGGCGGTGACCCAGAACTACGAGCGCTTACAGCCACTTCTG 3523  
Db 191 ArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeu 210  
Qy 3524 TGTACTTGGTGGCTCAACCGTCAAGTGGTGCACCTGTGAGTGGCGCGCAGCTGGCC 3583  
Db 211 CysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAla 230  
Qy 3584 CACTTCTGACCACTCCAGGCTCTGCTGGGCGCTGCTGGGCGCTCATGCACTGCTG 3643  
Db 231 HisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMetAlaLeu 250

Qy 3644 GGCTACCCACTGCCCGCCAGCCGCTGCTGGAGTGAACCCACTTGGACTTCTGCGCCTGCC 3703  
Db 251 GlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrProGlyProAla 270  
Qy 3704 CACAGTGACTTCTCCAGAGATGACAGCACTTCTGGCTGCTGCTGCTGCTGCTGCTGCTG 3763  
Db 271 HisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeuGlnThrTrp 290  
Qy 3764 CTGTGGCGCTCGGCCAAGCACTTCAACCGGCTCAAGAGAAAGATGAGCTCCAGCAGCT 3823  
Db 291 LeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysLysMetGlnProProAlaAla 310  
Qy 3824 GCAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3856  
Db 311 AlaValThrLeuHisGlyAlaHisGlyPhe 321  
RESULT 3  
AAW78415  
ID AAW78415 standard; Protein; 260 AA.  
XX  
AC AAW78415;  
XX  
XX  
DT 06-NOV-2001 (first entry)  
XX  
XX  
DE Human protein SEQ ID NO 1077.  
XX  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200157190-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 05-FEB-2001; 2001WO-US04098.  
XX  
PR 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
XX  
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Zhang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
DR WPI; 2001-476283/51.  
DR N-PSDB; AAK51548.  
XX  
XX  
PT Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX  
PS Claim 20; Page 3306; 6221pp; English.  
XX  
XX  
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

Sequence	260 AA;
SQ	

Alignment Scores:

Pred. No.:	2,17e-72	Length:	260
Score:	1112.00	Matches:	258
Percent Similarity:	29.77%	Conservative:	1
Best Local Similarity:	29.66%	Mismatches:	1
Query Match:	11.79%	Indels:	610
DB:	22	Gaps:	1

US-09-931-704-3 (1-5087) X AAM78415 (1-260)

Qy	1247	CTGTGTGCCACCTCACCCCTCACCCCAACCCAGCTGGGACAGACACCTGAGGGCTG	1306
Db	1	MetValSerHisProHisProProProSerProArgTpdGlyGlnThrProGluGlyLeu	20
Qy	1307	CCAGTGTCTCCCGTGTGGGCGCGCGCTCATGCTTCTCGTCCATCTCTGCCACA	1366
Db	21	ProAlaLeuSerProCysGlyProGlyProArgSerCysPheSerSerIleLeuProThr	40
Qy	1367	GGGACTCGTGGGGATGTTAGCGTGCCTGTGCACGCTGCTCTGGCACTCCCTGCAGTG	1426
Db	41	GlyAspSerTpdGlyMetLeuAlaCysLeuCysThrValLeuThrHisLeuProAlaVal	60
Qy	1427	CCAGTCTCAATCGACAGGGGACCCAGGCGCTGGGCCCTTCATCCAGAAAACTATGAC	1486
Db	61	ProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIleGlnIlysThrTyrAsp	80
Qy	1487	CTCACCGCTACCTGGAGCACCAACTCCGACGCTTGGCTGGGACCTATGTGAGTATCCAG	1546
Db	81	ProThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThrTyr	96
Qy	1547	CGTAGGAATCTGGGAGTGTGGGAGGAGTAGGAGTGTGGGAAAGACAGTCTTAACCGTGG	1606
Db	96		96
Qy	1607	AGGTTCTGGTAAATCATGGGTGAGGAGGGGCTCTTTGGCTCCCACTGCCCTGTC	1666
Db	96		96
Qy	1667	TGTTCTATCTCTGCGCTTCCCTCTTAGTGGCGCCCCCACTTCCCATCCCTGGCCCCA	1726
Db	96		96
Qy	1727	GGACTAGGATGTGGCAGGCTTCGACCCGCTTGGCCCATTTGCCCACTGCGCTGCCAG	1786
Db	96		96
Qy	1787	CCCAGCCGCCCTCCCCCTGGGGCGGGGAGTCTCTCTGTTTACACCGTGTGTG	1846
Db	96		96
Qy	1847	GTGTCTTTGCGCGGCGGGTTGGGTGGGGACAGAGGGGCCCCACCTCCCATGCGCTGCG	1906
Db	96		96
Qy	1907	TTCCAGTCGCGCTCTGCCCCCCAGACCTGGGGCCCTGCTGCTGTGAGCCCAAGGGGCTTCC	1966
Db	96		96
Qy	1967	TTCCGTCTGCGCTCTCCCATCTAGCTGGGCCCTCTAGGGGGGTTCATGGGGAAGGGGACT	2026
Db	96		96
Qy	2027	GTAGGGAACCCAGGCAGTAGTCGCAGGGGGTTTAGGGTGTGGATGAGGTTATGCTGTAA	2086
Db	96		96

```
Db 96 ----- 96
Qy 3227 TGAACAAAGACCTAGCATGAGACAGCGCCAGTGTTCAGGGGACACAAAAATAGAAAC 3286
Db 96 ----- 96
Qy 3287 TTTGGAGACAGTATCTCTTGGTGGTGAGCAGCGGCTGTGCCCTCTCCCTTCCCTCC 3346
Db 96 ----- 96
Qy 3347 ACCCTCTCTTTACACAGCTGAACCTACTGGGCCCCCTTTCAACAGCAGCAGCTTCAAC 3406
Db 97 -----LeuAsnTyrLeuGlyProPheAsnGluProAspPheAsn 110
Qy 3407 CTTCCCGCTGGGGCAGAGACTGTGCCAGGGCCACTGTTGACTTGGAGGTGGCGGA 3466
Db 111 ProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArg 130
Qy 3467 AGCTCAATGACAAACTCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGT 3526
Db 131 SerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCys 150
Qy 3527 TACTTGGCTGGCTCAACGCTCAGCTGCCACTGCTGAGCTGCGCGCAGCTGCGCCAC 3586
Db 151 TyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHis 170
Qy 3587 TTCTGCACAGCTCCAGGGCTGCTGGGCAGCATTGGGGGCTCATGGCAGCTCTGGGC 3646
Db 171 PheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGly 190
Qy 3647 TACCCACTGCCCGCTGCTGGGACTGAACCCACTTGGACACTCTGGCCCTGCCAC 3706
Db 191 TyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrProGlyProAlaHis 210
Qy 3707 AGTGACTTCTCCAGAGTGAACACTTCTGGCTGCTGAGAGCTGCGACCTGGCTG 3766
Db 211 SerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeu 230
Qy 3767 TGGCGCTCGGCCAAGCACTTCAACCGCTCAAGAGAGATGACGCTCCAGCAGCTGCA 3826
Db 231 TrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnProProAlaAlaAla 250
Qy 3827 GTACCTCTGCACCTGGGGCTCATGGCTTC 3856
Db 251 ValThrLeuHisLeuGlyAlaHisGlyPhe 260

RESULT 4
AA25831
ID AA25831 standard; Protein; 253 AA.
AC AA25831;
AL
DT 16-OCT-2001 (first entry)
DE
XX Human protein sequence SEQ ID NO:1346.
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;
KW antiagregant; haemostatic; vulnerary; antitumor; osteopathic; cytostatic;
KW dermatological; allergic; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopoenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
XX
```

```
OS Homo sapiens.
FN WO200153455-A2.
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000WO-US35017.
XX
PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX WPI: 2001-457603/49.
XX N-PSDB; AA999772.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX Claim 20; Page 278; 1217pp; English.
XX
XX AA999166 to AA999904 encode the human proteins given in AA25225 to
XX AA25963. The proteins can have activities based on the tissues and
XX cells they are expressed in, such as: antiinflammatory; antirheumatic;
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;
XX cardiovascular; antianemic; antiagregant; haemostatic; vulnerary;
XX antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;
XX antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX encoding them can be used in gene therapy, antisense therapy and vaccine
XX production. The proteins and polynucleotides are useful for screening for
XX agonists or antagonists of a protein and for the treatment and diagnosis
XX of disorders associated with the activity of a protein e.g. inflammation,
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX anaemia, platelet disorders, thrombocytopoenia, wounds, burns, ulcers,
XX osteoporosis, severe combined immunodeficiency, eczema, allergic
XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX neurological disorders.
XX Sequence 253 AA;
SQ

Alignment Scores:
Pred. No.: 9.64e-58 Length: 253
Score: 910.50 Matches: 231
Percent Similarity: 26.84% Conservative: 2
Best Local Similarity: 26.61% Mismatches: 14
Query Match: 9.65% Indels: 621
DB: Gaps: 4

US-09-931-704-3 (1-5087) x AA25831 (1-253)
Qy 1253 TCCACCTCACCCTCCACCCAGCCAGCCAGCTGGGGGACACACCTGAGGGGCTGCCAGCT 1312
Db 7 AlaArgProSerThrProProAlaSerGlyArgGlyAlaAlaPro---GlyArgPro--- 24
Qy 1313 GCTTCCCGCTGTGGCGCCCGCGGCTCATGCTTCTCGTCCATCTCTGCCACAGGGGAC 1372
Db 25 -----GlyProSerProMetAsp-----LeuArgAlaGlyAsp 35
Qy 1373 TCGTGGGGAGATTAGCTGTGTGCGAGCGGTCTGTGACCTCTCCCTGCACTGCCAGCT 1432
Db 36 SerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAla 55
Qy 1433 CTCATCCACAGGGGACCCCGGCTGGCCCTCCATCCAGAAACCTATGACCTCAC 1492
Db 56 LeuAsnArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThr 75
```



```

Db 186 LeuProGlnProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAsp 205
Qy 3713 TTCCTCAGAGAGTGGAGCTTCTGGCTGTGAAGAGCTGCAGACCTGGCTGGCGC 3772
Db 206 PheLeuGlnLysMetAspPheTrpLeuLysGluLeuGlnThrTrpLeuTrpArg 225
Qy 3773 TCGGCCAAGGACTTCAACCGCTCAAGAAGAGATGCAGCTCCAGAGCTGCAGTCAAC 3832
Db 226 SerAlaLysAspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThr 245
Qy 3833 CTGCACCTGGGGCTCATGGCTTC 3856
Db 246 LeuHisLeuGlyAlaHisGlyPhe 253

RESULT 5
ID AAB19586 standard; Protein; 215 AA.
XX AC AAB19586;
XX AC AAB19586;
XX DT 22-JAN-2001 (first entry)
XX XX Human interleukin-B60 (IL-B60).
XX KW Interleukin-B60; IL-B60; human; cytokine; cytokine-like factor-1;
XX KW haematopoietic; inflammation; antiinflammatory; autoimmune disease;
XX KW therapy.
XX OS Homo sapiens.
XX FH Key
XX FT Peptide
XX FT /label= Signal_peptide
XX FT Protein
XX FT /label= Mature-protein
XX FN WO200053631-A1.
XX PD 14-SEP-2000.
XX PF 09-MAR-2000; 2000WO-US06182.
XX PR 11-MAR-1999; 99US-0267901.
XX PA (SCHE ) SCHERING CORP.
XX PI Oppmann B, Timans JC, Kastelein RA, Bazan JF;
XX DR WPI; 2000-587426/55.
XX DR N-PSDB; AAA88546.
XX PT Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes, and for
XX PT polypeptides, and nucleic acids, useful in research, diagnosis and for
XX PT treating inflammatory and autoimmune disorders -
XX PS Claim 1; Page 15-16; 97pp; English.
XX CC The present sequence is that of human interleukin-B60 (IL-B60), a
XX CC novel, small soluble cytokine-like protein that exhibits structural
XX CC motifs characteristic of a member of the long-chain cytokines, and
XX CC which shows homology to granulocyte colony stimulating factor and
XX CC interleukin-6. IL-60B may have either stimulatory or inhibitory
XX CC effects on haematopoietic cells, including e.g. lymphoid cells,
XX CC such as T-cells, B-cells, natural killer cells, macrophages,
XX CC dendritic cells, haematopoietic progenitors, etc. Methods are
XX CC provided for modulating the physiology or development of a cell or
XX CC tissue culture cells by contacting the cell with an agonist or
XX CC antagonist of IL-B60 or an agonist of antagonist of a complex of
XX CC mature IL-B60 and its partner, cytokine-like factor-1 (CLF-1, see
XX CC AAB19588). The IL-B60/CLF-1 cytokine serves as a key physiological
XX CC factor in motor neuron development and regeneration. IL-60B, its
XX CC agonists and antagonists may be used to treat inflammatory or

```

```

CC autoimmune disorders and also for drug screening.
XX SQ Sequence 215 AA;
XX Alignment Scores:
Pred. No.: 1.03e-56 Length: 215
Score: 896.00 Matches: 170
Percent Similarity: 90.72% Conservative: 6
Best Local Similarity: 87.63% Mismatches: 6
Query Match: 9.50% Indels: 12
DB: Gaps: 2

US-09-931-704-3 (1-5087) x AAB19586 (1-215)
Qy 3311 GGTGAGCCAGCGGCTCTGCCCTCC-----TCCTTCCCACATCACC----- 3349
Db 22 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 41
Qy 3350 -----CTCTCTTTTCACAGCTGAACCTACCTGGGCGCCCTTTCAACGAG 3394
Db 42 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPheAsnGlu 61
Qy 3395 CCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGGCCACTGTTGACTTG 3454
Db 62 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 81
Qy 3455 GAGGTGTGGCAAGCTCAATGACAAACTGGGGCTGACCCAGAACTACGAGGCTACAGC 3514
Db 82 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 101
Qy 3515 CACCTTCTGTGTACTTGGTGGCTCAACCGTCAAGCTGAGCTGCCACTGTGAGCTGCCCGC 3574
Db 102 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgArg 121
Qy 3575 AGCTTGGCCCACTTCTGCACAGCTCCAGGGCTGCTGGGCGAGCATTCGCGCGCTCATG 3634
Db 122 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 141
Qy 3635 GCAGCTCTGGGCTACCCACTGCCCGCGCTGCTGGGACTGAACCCACTTGGACTCTCT 3694
Db 142 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpPro 161
Qy 3695 GGCCCTGCCCACTGAGTCTTCTCCAGAGATGGAGCAGCTTCTGGCTGTGAAGGCTG 3754
Db 162 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLysGluLeu 181
Qy 3755 CAGACTGGCTGTGGGCTCGGCCAAGGACTTCAACCGCTCAAGAAGAGATGAGCCT 3814
Db 182 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 201
Qy 3815 CCAGCAGCTGCAGTCAACCTGCACCTGGGGGCTCATGGCTTC 3856
Db 202 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 215

RESULT 6
AAW29715
ID AAW29715 standard; Protein; 225 AA.
XX AC AAW29715;
XX AC AAW29715;
XX DT 09-NOV-1998 (first entry)
XX XX Human neurotrophic factor NNT-1.
XX KW NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;
XX KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
XX KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
XX KW peripheral neuropathy; dystrophy; neural retina degeneration;
XX KW common variable immunodeficiency; CVID; selective IgA deficiency;
XX KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
XX KW therapy.
XX OS Homo sapiens.

```

FH Key Location/Qualifiers  
 FT Peptide 1..27  
 FT Protein /label= Sig\_peptide  
 FT 28..225  
 FT /label= Mat\_protein  
 XX WO9833922-A1.  
 PN 06-AUG-1998.  
 XX 02-FEB-1998; 98WO-US02363.  
 XX 30-JAN-1998; 98US-0016534.  
 PR 03-FEB-1997; 97US-0792019.  
 XX (AMGE-) AMGEN INC.  
 PA Chang M, Elliot GS, Sarmiento U, Senaldi G;  
 PI WPI; 1998-437475/37.  
 DR N-PSDB; AAV47510-11.  
 XX Newly isolated nucleic acid encoding human or murine neurotrophic factor NNT-1 - useful for treatment of neurological and immunological diseases or inflammation, also as vaccine adjuvant  
 PT Claim 12; Fig 3; 120pp; English.  
 XX This is the amino acid sequence of a novel neurotrophic factor, designated NNT-1, that is a growth factor for neurons and for B or T cells. It was deduced from isolated cDNA (see AAV47510) and genomic DNA (see AAV47511) clones. Vectors containing the cDNA or genomic DNA and host cells are provided for use in the production of NNT-1 polypeptides. These are used to treat: (i) neurological or immunological diseases, specifically Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and degeneration of the neural retina, or conditions characterised by T or B cell defects, e.g. common variable immunodeficiency (CVID), selective IgA deficiency, hypogammaglobulinaemia and X-linked agammaglobulinaemia (claimed), but many others disclosed; and (ii) inflammation. NNT-1 is also able to boost immunoreactivity and antibody production following vaccination, and, since it inhibits tumour necrosis factor production, it may also be useful for treating sepsis. In addition, cells that have been engineered to express NNT-1 can be implanted, or nucleic acids are delivered in gene therapy vectors.  
 XX Sequence 225 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1,05e-56 Length: 225  
 Score: 896.00 Matches: 170  
 Percent Similarity: 90.72% Conservative: 6  
 Best Local Similarity: 87.63% Mismatches: 6  
 Query Match: 9.50% Indels: 12  
 DB: 19 Gaps: 2  
 US-09-931-704-3 (1-5087) x AAW29715 (1-225)  
 QY 3311 GGTGAGCAGCGCTCTGGCTCC-----TCCTCCCATCACC----- 3349  
 Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51  
 QY 3350 -----CTCTCCCTTTTACAGCTGAACCTACTCTGGGCGCCCTTTTCAACGAG 3394  
 Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71  
 QY 3395. CCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTGACTTG 3454  
 Db 72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProAlaThrValAspLeu 91

QY 3455 GAGGTGTGGCGAAGCCTCAATGACAAACTGCGGCTGACCCAGCACTACGAGGCTACAGC 3514  
 Db 92 GluValTTPArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111  
 QY 3515 CACCTTCTGTGTTACTTGGTGGCCTCAACCGTCAAGGTGGCTGCCACTGCTGAGCTGCGCGC 3574  
 Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgArg 131  
 QY 3575 AGCCTGGCCCACTTCTGTCACAGCCTCCAGGGCCTGCTGGGCGAGCATTCGGGGCGTCATG 3634  
 Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151  
 QY 3635 GCAGCTCTGGGTACCCACTGCCAGCGCTGCTGGGACTGAACCCACTTGGACTCCT 3694  
 Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpPro 171  
 QY 3695 GGCCTGCCCACAGTACTTCTCCAGAGATGGAGCAGCTTCTGGCTGTGAAGGAGCTG 3754  
 Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLysGluLeu 191  
 QY 3755 CAGACCTGGCTGTGGCGCTCGGCCCAAGGACTTTCAACCGGCTCAAGAGAAAGATGCAGCCT 3814  
 Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211  
 QY 3815 CCAGCAGCTGCAGTCACCTGCACCTGGGGGCTCATGGCTTC 3856  
 Db 212 ProAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225  
 RESULT 7  
 AAW56141  
 ID AAW56141 standard; Protein; 225 AA.  
 XX AAW56141;  
 XX 13-JUL-1998 (first entry)  
 DT Amino acid sequence of human neurotrophic factor NNT-1.  
 DE Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;  
 KW treatment; neurological disease; degeneration; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.  
 XX Homo sapiens.  
 OS  
 Key Location/Qualifiers  
 FT Peptide 1..27  
 FT Protein /note= "signal peptide"  
 FT 28..225  
 FT /note= "mature protein"  
 XX US5741772-A.  
 PN 21-APR-1998.  
 PD 03-FEB-1997; 97US-0792019.  
 PF 03-FEB-1997; 97US-0792019.  
 PR (AMGE-) AMGEN INC.  
 PA Chang M;  
 PI WPI; 1998-260526/23.  
 DR N-PSDB; AAV22652.  
 XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -  
 PT useful for stimulating growth of motor and sympathetic neurons  
 XX Claim 1; Fig 3; 41pp; English.  
 XX The present sequence represents a human neurotrophic factor, designated  
 CC NNT-1, which is capable of stimulating growth of motor or sympathetic  
 CC neurons. The NNT-1 protein is useful in the treatment of neurological

CC diseases characterised by the degeneration and death of particular  
 CC classes of neurons. These diseases specifically include Parkinson's  
 CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,  
 CC stroke and various degenerative disorders affecting vision.

XX  
 SQ Sequence 225 AA;

Alignment Scores:  
 Pred. No.: 1,05e-56 Length: 225  
 Score: 896.00 Matches: 170  
 Percent Similarity: 90.72% Conservative: 6  
 Best Local Similarity: 87.63% Mismatches: 6  
 Query Match: 9.50% Indels: 12  
 DB: 19 Gaps: 2

US-09-931-704-3 (1-5087) x AAWS6141 (1-225)

QY 3311 GGTGAGCCAGCGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349  
 Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51  
 QY 3350 -----CTCTCCTTTTCACAGCTGAACCTACCTGGGCCCCCTTTCACAGAG 3394  
 Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPheAsnGlu 71  
 QY 3395 CCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGGCCACCTGTTGACTTG 3454  
 Db 72 ProaspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91  
 QY 3455 GAGGTGTGGCGAAGCTCAATGACAACTGGCGCTGACCCAGAACTACGAGSCCTACAGC 3514  
 Db 92 GluValTyrArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111  
 QY 3515 CACCTTCTGTATTCTTGGTGGCTCAACGCTCAGGCTGCCACTGCTGAGCTGGCGCGC 3574  
 Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131  
 QY 3575 AGCTGGCGCCACTTCTGACACAGCTCCAGGCGCTGCTGGGCGCATTCGGGCGCTCATG 3634  
 Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151  
 QY 3635 GCAGCTCTGGGTACCCACTGCCCGAGCGCTGCTGGGAGTGAACCCACTTGGACTCTCT 3694  
 Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171  
 QY 3695 GGCCCTGCCACAGTCACTTCTCAGAGATGGAGCTTCTGCTGCTGAGGAGCTG 3754  
 Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191  
 QY 3755 CAGACTGCTGCTGGCGCTCGGCCCAAGGACTTCAACCGGCTCAAGAAGATGAGCCT 3814  
 Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211  
 QY 3815 CCAGAGCTGCGAGTCACTGACCTGACCTGGGGCTCATGGCTTC 3856  
 Db 212 ProAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

# RESULT 8

AAW94466

ID AAW94466 standard; Protein; 225 AA.

XX

AC AAW94466;

XX

DT 22-APR-1999 (first entry)

XX Human cardiotrophin-like cytokine protein.

DE

XX Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;  
 KW CLC; IL-6; diagnosis; detection; immune system-related disorder;  
 KW cancer; cardiac disorder; heart failure; hypertension; cancer;  
 KW autoimmune disorder; infection.

XX

OS Homo sapiens.

XX Key Location/Qualifiers  
 FH Peptide 1..27  
 FT /label= signal  
 FT Protein 28..225  
 FT /label= Cardiotrophin-like\_cytokine  
 FT Domain 74..79  
 FT /label= CD-I  
 FT /note= "conserved domain"  
 FT Domain 150..156  
 FT /label= CD-II  
 FT /note= "conserved domain"  
 FT Domain 194..198  
 FT /label= CD-III  
 FT /note= "conserved domain"

XX W09900415-A1.

XX 07-JAN-1999.

XX 29-JUN-1998; 98WO-US13129.

XX 30-JUN-1997; 97US-0051311.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Shi Y;

XX WPI; 1999-095678/08.

XX DR N-PSDB; AAX16161.

XX New isolated cardiotrophin-like cytokine nucleic acid - used to

XX develop products for treating cardiac and immune system disorders,

XX e.g. heart failure, hypertension, cancers, autoimmune disorders and

XX infections

XX Claim 1; Fig 1; 103pp; English.

XX The present invention relates to a novel cardiotrophin-like cytokine

XX (CLC) protein which is a member of the interleukin 6 (IL-6) cytokine

XX family. The present sequence represents the human CLC protein. The

XX present invention also describes screening methods for identifying

XX agonists and antagonists of CLC activity, as well as methods for

XX detecting cardiac and immune system-related disorders and

XX therapeutic methods for treating cardiac and immune system-related

XX disorders, e.g. heart failure, hypertension, cancers, autoimmune

XX disorders and infections.

XX SQ Sequence 225 AA;

Alignment Scores:

Pred. No.: 1.05e-56 Length: 225

Score: 896.00 Matches: 170

Percent Similarity: 90.72% Conservative: 6

Best Local Similarity: 87.63% Mismatches: 6

Query Match: 9.50% Indels: 12

DB: 20 Gaps: 2

US-09-931-704-3 (1-5087) x AAWS6141 (1-225)

QY 3311 GGTGAGCCAGCGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349

Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51

QY 3350 -----CTCTCCTTTTCACAGCTGAACCTACCTGGGCCCCCTTTCACAGAG 3394

Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPheAsnGlu 71

QY 3395 CCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGGCCACCTGTTGACTTG 3454

Db 72 ProaspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91

QY 3455 GAGGTGTGGCGAAGCTCAATGACAACTGGCGCTGACCCAGAACTACGAGSCCTACAGC 3514

Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111  
 QY 3515 CACCTTCTGTGTTACTTGGCTCAACCGTCAAGGCTGCCACTGCTGAGTGGCGCGC 3574  
 Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131  
 QY 3575 AGCTGCGCCACTTCTGCACAGCTCCAGGCGCTGCTGGGCGCATTTGGCGGCTCATG 3634  
 Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151  
 QY 3635 GCAGCTTGGCTACCTACCTCCCGCGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGCT 3694  
 Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171  
 QY 3695 GGCCTCTCCACAGTCTCTCCAGAGATGAGCACTTCTGGCTGCTGCTGCTGCTGCTGCTG 3754  
 Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191  
 QY 3755 CAGACTTGGCTGGCGCTCGGCGCTCAAGGCTTCAACCGGCTCAAGAGATGCGAGCT 3814  
 Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211  
 QY 3815 CCAGCAGCTCAGTCACTCCCTGCACCTGGGCGCTCATGGCTTC 3856  
 Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225  
 RESULT 9  
 ID AAY87813 standard; Protein; 225 AA.  
 AC AAY87813;  
 DT 24-AUG-2000 (first entry)  
 DE Human NNT-1 protein.  
 KW NNT-1; human; neurotrophic factor; neurotrophic; neuroprotective; treatment;  
 KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;  
 KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
 KW Huntington's disease; peripheral neuropathy; neural retina degeneration;  
 KW retinopathy; immune disorder; hematopoietic disorder.  
 OS Homo sapiens.  
 XX  
 PN US6054294-A.  
 PD 25-APR-2000.  
 PF 12-DEC-1997; 97US-0988819.  
 PR 03-FEB-1997; 97US-0792019.  
 XX (AMGE-) AMGEN INC.  
 PA  
 PI Chang M;  
 XX  
 DR WPI; 2000-338492/29.  
 DR N-PSDB; AAA39481.  
 XX  
 PT New nucleic acids encoding neurotrophic factors useful for stimulating  
 PT growth of motor or sympathetic neurons for treating neuron cell damage  
 XX  
 PS Claim 1c; Fig 3; 42pp; English.  
 XX  
 CC This invention describes a novel nucleic acid molecule (I) encoding a  
 CC novel neurotrophic factor (NNT-1) (II) which has neurotrophic,  
 CC neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and  
 CC ophthalmological activity. (I) is useful for producing NNT-1  
 CC polypeptides which are useful for treating patients in whom various  
 CC cells of the central, autonomic, or peripheral nervous system have

CC degenerated and/or have been damaged by congenital disease, trauma,  
 CC mechanical damage, surgery, stroke, ischemia, infection, metabolic  
 CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1  
 CC proteins are used to treat diseases like Alzheimer's, Parkinson's,  
 CC amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's  
 CC disease, peripheral neuropathy induced by diabetes or other metabolic  
 CC disorders, and/or dystrophies or degeneration of the neural retina such  
 CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of  
 CC night blindness, progressive cone-rod degeneration, immune disorders and  
 CC hematopoietic disorders. (I) is effective in treating neurological  
 CC conditions and promotes neuron regeneration. Neural functions are  
 CC effectively restored in patients suffering from various neurological  
 CC disorders. This sequence represents the human NNT-1 protein described in  
 CC the method of the invention.  
 XX

SQ Sequence 225 AA;

#### Alignment Scores:

Pred. NO.: 1.05e-56 Length: 225  
 Score: 896.00 Matches: 170  
 Percent Similarity: 90.72% Conservative: 6  
 Best Local Similarity: 87.63% Mismatches: 12  
 Query Match: 9.50% Indels: 2  
 DB: 21 Gaps: 2

US-09-931-704-3 (1-5087) x AAY87813 (1-225)

QY 3311 GGTGAGCCAGCGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349  
 Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51  
 QY 3350 -----CTCTCTCTTTTTCACAGCTGAACTACTGCTGGCGCCCTTTTCAACGAG 3394  
 Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71  
 QY 3395 CCAGCTTCAACCTTCCCGCTGGGGCGAGAGACTCTGCCAGGCGCCTGTTGACTTG 3454  
 Db 72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91  
 QY 3455 GAGGTGTGGCAAGCTCAATGACAACTGCGGCTGACCCAGCACTACGAGGCTACAGC 3514  
 Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111  
 QY 3515 CACCTTCTGTGTTACTTGGCTCAACCGTCAAGGCTGCTGCTGCTGCTGCTGCTGCTG 3574  
 Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131  
 QY 3575 AGCTGCGCCACTTCTGCACAGCTCCAGGCGCTGCTGGGCGCATTTGGCGGCTCATG 3634  
 Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151  
 QY 3635 GCAGCTCTGGGCTACCCACTGCGCGCTGCTGGGCGCTGCTGGGCGCTGCTGGGCGCT 3694  
 Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171  
 QY 3695 GGCCTCTCCACAGTCTCTCCAGAGATGAGCACTTCTGGCTGCTGCTGCTGCTGCTGCTG 3754  
 Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191  
 QY 3755 CAGACTTGGCTGGCGCTCGGCGCTCAAGGCTTCAACCGGCTCAAGAGATGCGAGCT 3814  
 Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211  
 QY 3815 CCAGCAGCTCAGTCACTCCCTGCACCTGGGCGCTCATGGCTTC 3856  
 Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 10

AAG63543

ID AAG63543 standard; Protein; 225 AA.

XX

AC AAG63543;

XX





DR WPI; 2002-280867/32.  
DR N-PSDB; ABK11647.  
XX  
PT Treating Immunoglobulin E-related disease, modulating IgE levels in a  
PT patient, preventing IgE-related disease and treating allergic diseases,  
PT involves administering NNT-1 inhibitor to a patient -  
XX  
PS Claim 2; Fig 3; 63pp; English.  
XX  
XX The invention relates to treating Immunoglobulin E (IgE)-related disease,  
CC modulating IgE levels in a patient, preventing an IgE-related disease,  
CC and treating allergic diseases, comprising administering a  
CC therapeutically effective amount of novel neurotrophic factor (NNT)-1  
CC inhibitor to a patient. Also included are a method of diagnosing an  
CC IgE-related disease or susceptibility to an IgE-related disease, by  
CC determining the presence or amount of expression of an NNT1 polypeptide  
CC encoded by a NNT1 nucleotide sequence, its fragment or naturally  
CC occurring variant, and diagnosing an IgE-related disease or  
CC susceptibility of an IgE-related disease based on the presence or amount  
CC of expression of the polypeptide and a pharmaceutical composition for use  
CC in treating IgE-related disease, comprising the NNT1 inhibitor.  
CC The NNT1 inhibitor is useful for preventing and treating IgE-related  
CC disease, modulating IgE levels, and treating allergic diseases e.g.  
CC Type I allergic disease, allergic rhinitis, eczema, dermatitis,  
CC pollinosis, asthma, immune diseases and disorders, diseases involving  
CC abnormal cell proliferation including cancer, arteriosclerosis and  
CC vascular restenosis, diseases and conditions relating to dysfunction of  
CC immune system including rheumatoid arthritis, psoriatic arthritis,  
CC inflammatory arthritis, osteoarthritis, inflammatory joint disease,  
CC autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory  
CC bowel disease, transplant rejection, and graft versus host disease, and  
CC reproductive diseases and disorders including infertility, miscarriage,  
CC preterm labour and delivery, and endometriosis. The present sequence  
CC represents human NNT1.  
XX  
SQ Sequence 225 AA;

Alignment Scores:  
Pred. No.: 1,05e-56 Length: 225  
Score: 896.00 Matches: 170  
Percent Similarity: 90.72% Conservative: 6  
Best Local Similarity: 87.63% Mismatches: 6  
Query Match: 9.50% Indels: 12  
DB: 23 Gaps: 2

US-09-931-704-3 (1-5087) x AAU78176 (1-225)

QY 3311 GGTGACGCGGCTCTGCGCTCC-----TCCTTCCCATCACC----- 3349  
Db |||:||||| ||||| :|||: |||:  
32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51  
QY 3350 -----CTCTCTTTTTCACAGCTGAACCTACCTGGGCCCTTTTCAACGAG 3394  
Db |||:||||| ||||| :|||: |||:  
52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71  
QY 3395 CAGACTTCAACCTCCCGCTGGGGGAGAGACTTGCACGGGCCACTGTGACTTG 3454  
Db |||:||||| ||||| :|||: |||:  
72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91  
QY 3455 GAGGTGTGGAGCTCAATGACAACTCGCGCTGACCCAGCACTACGAGGCTACGC 3514  
Db |||:||||| ||||| :|||: |||:  
92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111  
QY 3515 CACCTTCTGTGTACTTGGTGGCCCTCAACCGTCAGGCTGCCACTGTGAGCTGCGCCG 3574  
Db |||:||||| ||||| :|||: |||:  
112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgArg 131  
QY 3575 AGCTGGCCCACTTCTGACACGACCTCCAGGCGCTGTGCGGCGCATGCGGCGTCAATG 3634  
Db |||:||||| ||||| :|||: |||:  
132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151  
QY 3635 GCAGCTCTGGGCTTACCACTGCCCGGAGCGGCTGCTGGGACTGAACCACTTGACCTCT 3694  
Db |||:||||| ||||| :|||: |||:

Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171  
QY 3695 GGCCTGCCCCAGTACTTCTCCAGAGATGGACGACTTCTGGCTGCTGAAGGAGCTG 3754  
Db |||:||||| ||||| :|||: |||:  
172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191  
QY 3755 CAGACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGACGCT 3814  
Db |||:||||| ||||| :|||: |||:  
192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211  
QY 3815 CCAGCAGCTGAGTACCTGCACCTGGGGGCTCATGGCTTC 3856  
Db |||:||||| ||||| :|||: |||:  
212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225  
RESULT 12  
ABB40317  
ID ABB40317 standard; Peptide; 164 AA.  
XX  
AC ABB40317;  
XX  
DT 04-FEB-2002 (first entry)  
XX  
DE Peptide #7823 encoded by human foetal liver single exon probe.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX Homo sapiens.  
XX WO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US006659.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024283.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.

Human genome-derived single exon nucleic acid probes useful for  
analyzing gene expression in human fetal liver -

Claim 27; SEQ ID NO 32952; 639pp + sequence listing; English.

The invention relates to a single exon nucleic acid probe for  
measuring human gene expression in a sample derived from human foetal  
liver. The single exon nucleic acid probes may be used for predicting,  
measuring and displaying gene expression in samples derived from human  
fetal liver. The present sequence is a peptide encoded by a single exon  
nucleic acid probe of the invention.  
Note: The sequence data for this patent did not form part of the  
printed specification, but was obtained in electronic format directly  
from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 164 AA;

Alignment Scores:  
Pred. No.: 5.91e-56 Length: 164  
Score: 885.00 Matches: 163  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.39% Mismatches: 0  
Query Match: 9.38% Indels: 0  
DB: 22 Gaps: 0

US-09-931-704-3 (1-5087) x ABB40317 (1-164)

QY 3365 CTGAATACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGGCA 3424

Db 1 LeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGlyAla 20

QY 3425 GAGACTGCCCCAGGCGCACTGTGACTGGAGGTGGGAGGCTCAATGCAAACTG 3484

Db 21 LysThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLysLeu 40

QY 3485 CGGCTGACCCAGAACTACAGCGCTACAGCCACTTCTGTGTTACTTGGTGGCTCCAC 3544

Db 41 ArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsn 60

QY 3545 CGTACAGGCTGCCACTGCTGAGCTGCGCGCAGCTTGGCCCTTCTGACACGCTCCAG 3604

Db 61 ArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeuGln 80

QY 3605 GGCCTGTGGGCGACATTTGGGGCGTCAATGCGAGCTCTGGGCTACCCACTGCCCCAGCCG 3664

Db 81 GlyLeuLeuGlySerIleAlaGlyValMetAlaLeuGlyTyrProLeuProGlnPro 100

QY 3665 CTGCTGGGAGCTGAACCCACTTGGACTCTGGCCCTGCGCCCTCAAGTCTCTCCAGAG 3724

Db 101 LeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGlnLys 120

QY 3725 ATGACGACGCTTCTGGCTGTGAAGAGCTGCAGACCTGGCTGGGCTACCCACTGCCCCAGCCG 3784

Db 121 MetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLysAsp 140

QY 3785 TTCAACCGGCTCAAGAGAGATGACAGCTCCAGAGTGCAGTCCACTGCGCTGGGG 3844

Db 141 PheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeuGly 160

QY 3845 GCTCATGGCTTC 3856

Db 161 AlaHisGlyPhe 164

RESULT 13

ABB24716

ID ABB24716 standard; Protein; 164 AA.

XX AC ABB24716;

DT 23-JAN-2002 (first entry)

DE Protein #6715 encoded by probe for measuring heart cell gene expression.

XX KW Human; gene expression; heart; microarray; vascular system;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease.

OS Homo sapiens.

XX WO200157274-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00666.

XX PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

DR

XX Single exon nucleic acid probes for analyzing gene expression in human hearts -

PT Claim 15; SEQ ID No 26486; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21535-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, the human heart and vascular system e.g. cardiovascular disease, diagnosing, grading, staging, monitoring and prognosing diseases of the hypertension, cardiac arrhythmias and congenital heart disease. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 164 AA;

QY Alignment Scores:

Pred. No.: 5,91e-56 Length: 164

Score: 885.00 Matches: 163

Percent Similarity: 100.00% Conservative: 1

Best Local Similarity: 99.39% Mismatches: 0

Query Match: 9.38% Indels: 0

DB: 22 Gaps: 0

US-09-931-704-3 (1-5087) x ABB24716 (1-164)

QY 3365 CTGAATACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGGCA 3424

Db 1 LeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGlyAla 20

QY 3425 GAGACTTGGCCAGGCGCACTGTGACTTGGAGGTGGGAGGCTCAATGCAAACTG 3484

Db 21 LysThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLysLeu 40

QY 3485 CGGCTGACCCAGAACTACAGCGCTACAGCCACTTCTGTGTTACTTGGTGGCTCCAC 3544

Db 41 ArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsn 60

QY 3545 CGTACAGGCTGCCACTGCTGAGCTGCGCGCAGCTTGGCCCTTCTGACACGCTCCAG 3604

Db 61 ArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeuGln 80

QY 3605 GGCCTGTGGGCGACATTTGGGGCGTCAATGCGAGCTCTGGGCTACCCACTGCCCCAGCCG 3664

Db 81 GlyLeuLeuGlySerIleAlaGlyValMetAlaLeuGlyTyrProLeuProGlnPro 100

QY 3665 CTGCTGGGAGCTGAACCCACTTGGACTCTGGCCCTGCGCCCTCAAGTCTCTCCAGAG 3724

Db 101 LeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGlnLys 120

QY 3725 ATGACGACGCTTCTGGCTGTGAAGAGCTGCAGACCTGGCTGGGCTACCCACTGCCCCAGCCG 3784

Db 121 MetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLysAsp 140

QY 3785 TTCAACCGGCTCAAGAGAGATGACAGCTCCAGAGTGCAGTCCACTGCGCTGGGG 3844

Db 141 PheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeuGly 160

QY 3845 GCTCATGGCTTC 3856

Db 161 AlaHisGlyPhe 164

RESULT 14

AM61118

ID AM61118 standard; Protein; 164 AA.

XX AC AM61118;

XX



DB:	*	22	Gaps:	0
US-09-931-704-3	(1-5087) x AAM73827	(1-164)		
QY	3365	CTGAAC <sup>T</sup> ACCTGGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGGGCA	3424	
Db	1	LeuAsn <sup>Tyr</sup> LeuGly <sup>Pro</sup> ProPheIa <sup>n</sup> Glu <sup>Pro</sup> AspPheAsn <sup>Pro</sup> ProIa <sup>n</sup> ArgLeuGly <sup>Ala</sup>	20	
QY	3425	GAGACT <sup>T</sup> CGCCAGGGCCACTGTTCACTTGGAGGTGTGGCGAAGCCTCAATGACAAACTG	3484	
Db	21	Lys <sup>Thr</sup> Leu <sup>Pro</sup> Arg <sup>Ala</sup> Thr <sup>Val</sup> Asp <sup>Leu</sup> Glu <sup>Val</sup> Trip <sup>Arg</sup> Ser <sup>Leu</sup> Asn <sup>Asp</sup> Lys <sup>Leu</sup>	40	
QY	3485	CGGCTGACCCAGAACTACGAGCCCTACAGCCACCTCTCTGTACTTGTGGTGGCTTCAAC	3544	
Db	41	Arg <sup>Leu</sup> Thr <sup>Gln</sup> Asn <sup>Tyr</sup> Glu <sup>Ala</sup> <sup>Tyr</sup> Ser <sup>His</sup> Leu <sup>Leu</sup> Cys <sup>Tyr</sup> Leu <sup>Arg</sup> Gly <sup>Leu</sup> Asn	60	
QY	3545	CGTCAGGCTGCCACTGCTAGCTGGCGCGCAGCTGGCCACTTGTGACACGACCTCCAG	3604	
Db	61	Arg <sup>Gln</sup> Ala <sup>Ala</sup> Thr <sup>Ala</sup> Glu <sup>Leu</sup> Arg <sup>Ser</sup> Leu <sup>Ala</sup> His <sup>Phe</sup> Cys <sup>Thr</sup> Ser <sup>Leu</sup> Gln	80	
QY	3605	GGCTGTCTGGGACGACTTGCGGCGCTCATGGCAGCTCTGGGCTACCCACTGCCGCCAGCCG	3664	
Db	81	Gly <sup>Leu</sup> Leu <sup>Gly</sup> Ser <sup>Ile</sup> Ala <sup>Gly</sup> Val <sup>Met</sup> Ala <sup>Leu</sup> Gly <sup>Tyr</sup> Pro <sup>Leu</sup> Pro <sup>Gln</sup> Pro	100	
QY	3665	CTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCACAGTGACTCTCTCCAGAAG	3724	
Db	101	Leu <sup>Pro</sup> Gly <sup>Thr</sup> Glu <sup>Pro</sup> Thr <sup>Trip</sup> Thr <sup>Pro</sup> Gly <sup>Pro</sup> Ala <sup>His</sup> Ser <sup>Asp</sup> Phe <sup>Leu</sup> Gln <sup>Lys</sup>	120	
QY	3725	ATGGACGACTCTGGCTGCTGAAGAGCTGCAGACTGGCTGTGGCGCTCGGCCAAGGAC	3784	
Db	121	Met <sup>Asp</sup> Asp <sup>Phe</sup> Thr <sup>Leu</sup> Leu <sup>Lys</sup> Glu <sup>Leu</sup> Gln <sup>Thr</sup> Trip <sup>Leu</sup> Thr <sup>Arg</sup> Ser <sup>Ala</sup> Lys <sup>Asp</sup>	140	
QY	3785	TTCAACCGGCTCAAGAAGAATGACGCTCCAGCAGCTGCAGTCCACCTGCACCTGGGG	3844	
Db	141	Phe <sup>Asn</sup> Arg <sup>Leu</sup> Lys <sup>Lys</sup> Lys <sup>Met</sup> Gln <sup>Pro</sup> Pro <sup>Ala</sup> Ala <sup>Val</sup> Thr <sup>Leu</sup> His <sup>Leu</sup> Gly	160	
QY	3845	GCTCATGGCTTC	3856	
Db	161	Ala <sup>His</sup> Gly <sup>Phe</sup>	164	

Search completed: January 27, 2003, 15:59:49  
Job time : 235.014 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 16:20:23 ; Search time 36.4279 Seconds  
(without alignments)  
5635.701 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 9432

Sequence: 1 aaactgcgagtgccctggc.....ccttgtaagtctctca 5087

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 244452

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2\_1/USFTO.spool/US09931704/runat\_27012003\_154129\_3760/app\_query.fasta\_1.7189  
-DB=Published Applications AA -QFMT=fastan -SURFIX=rapb -MINMATCH=0.1  
-LOOFCU=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09931704 @CGN 1.1 51 @runat\_27012003\_154129\_3760  
-NCPU=6 -ICPU=3 -NO\_XLPV -NO\_MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pcp:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pcp:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pcp:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pcp:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pcp:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pcp:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pcp:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pcp:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pcp:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pcp:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pcp:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pcp:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pcp:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	896	9.5	225	10	US-09-931-704-2 Sequence 2, Appli
2	885	9.4	164	10	US-09-864-761-40014 Sequence 40014, A
3	863	9.1	225	10	US-09-931-704-5 Sequence 5, Appli
4	281.5	3.0	1806	10	US-09-919-497-56 Sequence 56, Appli

5	267	2.8	1274	9	US-10-020-215-2	Sequence 2, Appli
6	259	2.8	714	10	US-09-861-597-10	Sequence 10, Appli
7	239.5	2.6	595	9	US-09-854-133-187	Sequence 187, App
8	239.5	2.6	595	10	US-09-738-973-187	Sequence 187, App
9	239.5	2.5	5179	9	US-10-025-380-1068	Sequence 1068, Ap
10	239.5	2.5	5179	10	US-09-922-217-1068	Sequence 1068, Ap
11	239.5	2.5	5179	10	US-09-833-263-1068	Sequence 1068, Ap
12	233	2.5	1567	10	US-09-835-232-2	Sequence 2, Appli
13	233	2.5	1806	10	US-09-919-497-56	Sequence 56, Appli
14	230.5	2.4	2005	10	US-09-735-367B-3	Sequence 3, Appli
15	226	2.4	2063	10	US-09-735-367B-2	Sequence 2, Appli
16	223.5	2.4	1690	10	US-09-788-043C-5	Sequence 5, Appli
17	222.5	2.4	877	12	US-10-041-770-2	Sequence 2, Appli
18	218.5	2.3	4019	9	US-09-854-133-425	Sequence 425, App
19	218.5	2.3	4019	10	US-09-738-973-425	Sequence 425, App
20	216.5	2.3	507	9	US-10-078-547-24	Sequence 24, Appli
21	216	2.3	1367	10	US-09-801-368-108	Sequence 108, App
22	215	2.3	802	10	US-09-823-240-2	Sequence 2, Appli
23	214	2.3	503	9	US-10-078-547-2	Sequence 2, Appli
24	213.5	2.3	529	10	US-09-861-597-2	Sequence 7, Appli
25	212.5	2.3	1317	10	US-09-963-896-7	Sequence 53, Appli
26	204.5	2.2	646	9	US-09-964-899-53	Sequence 4, Appli
27	203.5	2.2	606	10	US-09-861-597-4	Sequence 17, Appli
28	198.5	2.1	645	9	US-09-964-899-17	Sequence 15, Appli
29	198	2.1	593	9	US-09-964-899-15	Sequence 6, Appli
30	197.5	2.1	1234	10	US-09-854-173A-12	Sequence 12, Appli
31	197.5	2.1	1234	10	US-09-854-173A-12	Sequence 10, Appli
32	196.5	2.1	1056	9	US-09-861-597-1	Sequence 1, Appli
33	195.5	2.1	651	10	US-10-016-283-34	Sequence 34, Appli
34	195	2.1	1940	9	US-10-001-873-50	Sequence 50, Appli
35	193.5	2.1	1134	9	US-09-861-597-8	Sequence 8, Appli
36	192.5	2.1	606	10	US-09-894-998-47	Sequence 47, Appli
37	192.5	2.1	826	9	US-10-001-887-108	Sequence 2, Appli
38	191	2.0	638	10	US-09-858-664A-2	Sequence 5, Appli
39	190.5	2.0	1665	10	US-10-077-130-2	Sequence 5, Appli
40	190.5	2.0	2630	9	US-10-077-130-5	Sequence 17, Appli
41	190.5	2.0	7968	9	US-10-086-464-17	Sequence 2, Appli
42	190	2.0	731	9	US-10-078-547-2	Sequence 1, Appli
43	188.5	2.0	503	9	US-09-963-875-1	Sequence 10, Appli
44	188.5	2.0	1618	9	US-10-109-886-10	
45	188.5	2.0	2442	12	US-10-109-886-10	

#### ALIGNMENTS

RESULT 1

US-09-931-704-2  
; Sequence 2, Application US/09931704  
; Patent No. US20020041873A1  
; GENERAL INFORMATION:  
; APPLICANT: Senaldi, Giorgio  
; TITLE OF INVENTION: Methods and Compositions for Treating Ige-Related Disease Using Inhibitors  
; FILE REFERENCE: A-695  
; CURRENT APPLICATION NUMBER: US/09/931,704  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: US 60/236,436  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 2  
; LENGTH: 225  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-931-704-2

Alignment Scores:  
Pred. No.: 4.84e-43  
Score: 896.00  
Length: 225  
Matches: 170  
Percent Similarity: 90.72%  
Conservative: 6  
Best Local Similarity: 87.63%  
Mismatches: 6  
Query Match: 9.50%  
Indels: 12  
DB: 10  
Gaps: 2

```
US-09-931-704-3 (1-5087) x US-09-931-704-2 (1-225)
QY 3311 GGTGAGCCAGCGCTCTCCCTCC-----TCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCTTTTTCACAGCTGAACACTCTGGGCCCCCTTTCAACGAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCAACCTCCCGCTCGGGGCGAGAGACTCTGCCAGGGCCACTGTGTGACTTG 3454
Db 72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91
QY 3455 GAGGTGTGGCAGAGCTCAATGACAACTGCGGCTGACCCAGACTACGAGCCTACAGC 3514
Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGGTGGCTCAACCGTCAAGCTGCGGCTGCACTGCTGAGCTGCGCCG 3574
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeuArg 131
QY 3575 AGCTTGGCCCACTTTCACACAGCTCCAGGCGCTGCTGGGAGCAATTCGGGCGCTCATG 3634
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 CGAGCTGGGCTACCCACTGCGCCAGCGCTGCGCTGCGGAGCAATTCGGGCGCTCATG 3694
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171
QY 3695 GGCCTTGGCCACAGTACTTCTCCAGAGATGGAGCACTTCTGGCTGCTGAAGGAGCTG 3754
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLysGluLeu 191
QY 3755 CAGACTGGCTGTGGCTGCGCCCAAGAGCTTCAACCGCTCAAGAGAGATGCGAGCCT 3814
Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211
QY 3815 CCAGCAGCTGCAGTCACTGCTGCGCTGCGGCTCATGGCTTC 3856
Db 212 ProAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 2
US-09-864-761-40014
; Sequence 40014, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40014
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EST HUMAN HIT: A1752561.1, EVALUE 3.00e-66
; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUE 8.00e-03
US-09-864-761-40014

Alignment Scores:
Pred. No.: 1,96e-42 Length: 164
Score: 885.00 Matches: 163
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.39% Mismatches: 0
Query Match: 9.38% Indels: 0
DB: 10 Gaps: 0

US-09-931-704-3 (1-5087) x US-09-864-761-40014 (1-164)
QY 3365 CTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTTCCCGCTGGGGCA 3424
Db 1 LeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGlyAla 20
QY 3425 GAGACTCTGCCAGGGCCACTGTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAACTG 3484
Db 21 LysThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLysLeu 40
QY 3485 CGGCTGACCCAGAACTACGAGGCTACGCCACTTCTGTCTTACTTGGTGGCTTCAAC 3544
Db 41 ArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsn 60
QY 3545 CGTCAGCTGCCTGCTGAGCTGCGCGCAGCTGCGGCCACTTCTGCACACGCTTCCAG 3604
Db 61 ArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeuGln 80
QY 3605 GGCCTGCTGGGCGAGCATTTGCGGGCGTTCATGGCAGCTTGGGCTACCCACTGCCAGCCG 3664
Db 81 GlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGlnPro 100
QY 3665 CTGCTGGGAGTGAACCCACTTGGACTCTGCGCTGCCCGCCACAGTCACTTCTCCAGAG 3724
Db 101 LeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGlnLys 120
```







QY 2713 GAAC TAGAGGCGCAGTGGAGTGAAGCAAAAGGTTTCTGCAACAAGGAACAAGAGAGAGAT 2654  
Db 1210 -----GlyGluAsnGly 1213  
QY 2653 GAGAGTGGCCATTGGAACAGAGCCCATTTGCCGCTGACACACCCAGCCGACCCAGCT 2594  
Db 1214 AsnValGlyProGlnGlyPro-----ProGlyProProGly 1225  
QY 2593 GAGGGCTGGGAGAGACGGGTACTCACTAACACTCCGAAAGAAGGTTTCAGCTGATAGG 2534  
Db 1225 ----- 1225  
QY 2533 AGCAGGTTGCAGAGGAAGAATAAGGAGGAGTCTCATGAGTAGGACAGAGGAAGGCTA 2474  
Db 1226 ProArgGlyProGlnGlyProAsnGly-AlaAspGlyProGlnGlyProProGlySerVa 1245  
QY 2473 TGGAGACAGCATGTATAGGAGAGGCGAGAGCCCTTCCAAACCTCGGGGAGGGCG 2414  
Db 1245 lGlySerValGlyGlyValGlyGlyGlyGlyGlyPro-----GlyGluAl 1260  
QY 2413 TGGGGCCAGGGAAGGGATAGGGAGATGTATAACTGGGAAGAAGGAGAGCAAGGG 2354  
Db 1260 agly----- 1261  
QY 2353 GAGAGCCTGGGATAGGCAAGGCGCCAGGGGAGGTTTCACACGCTGACACAGCAGAGAC 2294  
Db 1261 ----- 1261  
QY 2293 AAAC TAATTCCTCGGTGTATCTCTGGGGTGGAGCAGCTCCCGCCGACCTCTTCCCT 2234  
Db 1262 -----AsnProGlyProProGlyGluAlaGlyVa 1271  
QY 2233 TTTCTTTCCCTCCAGTAGGTCTCTGGGATGTGGAGAGAGAAACACGGGAGAGGG 2174  
Db 1271 lGlyGlyPro-----LysGlyGluArgGlyGluGlySGL 1282  
QY 2173 GCGGTT---CCCATGTGCTCGCTCGCTCTCCAAACCTCTCTCTCTCTCTCTCTCTCT 2117  
Db 1282 yGluAlaGlyProProGlyAlaAlaGlyPro-----ProGlyAlaLysGlyPr 1298  
QY 2116 TCTGAACACCTTGACACACCCCAATCTTACACATTAACCTTCCATCCACACCTTAA 2057  
Db 1298 oProGlyAspAspGlyProLysGlyAsnProGlyProVal----- 1311  
QY 2056 CCCCCTGCCACTACTGCTGGGTCTCTACAGTCCCTCTCCCTCCCTCCCTCCCTCCCT 2014  
Db 1312 -----GlyPheProGlyAspProGlyProGlyGluLeuGlyPr 1325  
QY 2013 -----CATGACCCC-----CCTAG 2000  
Db 1325 oAlaGlyGlnAspGlyValGlyGlyAspLysGlyGluAspGlyAspProGlyGlnPro-G 1345  
QY 1999 GAGGCCAGCTAGTAGGAGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1947  
Db 1345 lyProGlyProSerGlyGlyAlaGlyProProGlyProProGlyProGlyProGlyPro 1365  
QY 1946 -----GCAGCAGGCCCCAG---GTCTGGGGGAGAGGAGGAGGAGGAGGAGGAGGAG 1895  
Db 1365 roGlyAlaAlaGlyAlaGlyGlyArgGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1385  
QY 1894 AGGTGGGCCCCCTCTGTCCC-----CACCCCAACCCCGCCCGCCGCAAG 1853  
Db 1385 laGluGlyPro-----ProGlyLysThrGlyProValGlyProGlnGlyProAlaGlyL 1403  
QY 1852 AGACACCAACACGCTGTAACAGAGGAGACTTCCCGGCCCCCAGAGGAGGAGGAGGAGGAG 1794  
Db 1403 yProGlyProGluGlyLeu---ArgGly---IleProGlyProValGlyGluGlnGlyL 1421  
QY 1793 -----GGCTGGGCTGCAGCCAGTGGGGCAATGGCGCAAGGCGGCTCGAGGCGCTGCCCA 1739  
Db 1421 euProGlyAlaAlaGlyGlnAspGlyProProGlyProMetGlyProProGlyLeuProG 1441

QY 1738 C-----ATGCCCTAGTCTCTGGGCGCCAGGGATGGGGAAGTGGGG----- 1702  
Db 1441 lyLeuLysGlyAspProGlySerLys---GlyGluLysGlyHisProGlyLeuIleGlyL 1460  
QY 1701 -----GGGCCACCTTAAGAGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1666  
Db 1460 eulleGlyProProGlyGluGlnGlyGluLysGlyAspArgGlyLeuProGlyThrGlnG 1480  
QY 1665 -----ACAGGGGACTGGTGGGAGCCCAAGAGAGGCGCCCTCTCTCCACCC 1626  
Db 1480 lySerProGlyAlaLysGlyAspGlyGlyIleProGlyProAlaGlyProLeuGlyPro 1500  
QY 1625 CATCATTTACCAGAACCTCCACGGTTAGGACTGTCTTTCCCAACTCTCTCTCTCTCC 1566  
Db 1500 roGly----- 1501  
QY 1565 CAATCCAGATTCCTACGCTGGATACTCATAGGTCCCAAGCTGCGGAGTTGCT 1506  
Db 1502 -----ProProGlyLeu----- 1505  
QY 1505 GCTCCAGGTAGCGGTGAGTGCATAGGTTTCTGGATGGAGGGGCGAGCCCTGGGTCC 1446  
Db 1506 -----ProGlyProGlnGlyP 1511  
QY 1445 CTGTGCGATTGAGAGCTGGGCACTGCAGGGAGGAGTGCAGAGCACCGTGCACAGCAGCTA 1386  
Db 1511 roLysGlyAsnLysGlySerThr----- 1518  
QY 1385 ACATCCCCCAGAGTCCCTCTGTGGGAGGATGGACGAGAGCATGACGCGGCG-----CCGG 1329  
Db 1519 -----GlyProAlaGly-----GlnLysGlyAspSerGlyLeuProG 1531  
QY 1328 GCCCACACGGGGAAGCAGCTGGCAGCCCTCAGGTGTCTGTCCCACTCTGGGTGGGTG 1269  
Db 1531 lyProProGly---ProProGlyProProGlyGluValleGlnProLeuProIleLeuS 1550  
QY 1268 GGGGCTCAGGCTGGGACACACAGTGGGAGGAGGAGGAGGAGT-----GAGGAGG 1221  
Db 1550 er-----SerLysLysThrArgHisThrGluGlyMetGlnAlaAspAlaAspAspA 1568  
QY 1220 GGAACACTCTGCGCAAGACGGGAGAGAA 1192  
Db 1568 snlleLeuAspTyrSerAspGlyMetGlu 1577  
RESULT 5  
US-10-020-215-2  
; Sequence 2, Application US/10020215  
; Publication No. US2003008347A1  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY  
; APPLICANT: PELES, EIOR  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF ALP RELATED DISORDERS  
; FILE REFERENCE: 038602/1290  
; CURRENT APPLICATION NUMBER: US/10/020,215  
; CURRENT FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: 09/095,443  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/049,477  
; PRIOR FILING DATE: 1997-06-11  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1274  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: ALP  
; OTHER INFORMATION: polypeptide sequence  
US-10-020-215-2  
Alignment Scores: 5.99e-08 Length: 1274  
Pred. No.: 267.00 Matches: 288  
Score:

Percent Similarity:	30.94%	Conservative:	112
Best Local Similarity:	22.27%	Mismatches:	394
Query Match:	2.83%	Indels:	499
DB:	9	Gaps:	73

US-09-931-704-3 (1-5087) x US-10-020-215-2 (1-1274)	
QY 7	CGAGTGGCGCTGGCGATGGGATTATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCA 66
Db	
342 ArgGlnGlnLeuLeuAspArgGluLeuLysLysLysProProArgProThrAlaPro 361	
QY 67	-----CTCGCGCAGCGCTCGGGAGAGAGCGCGACCCGGCGCGGCCAGCCGACGCC 117
Db	
362 LysProLeuLeuProArgGluGluSerGluAlaValGluAlaGlyAspProGlu 381	
QY 118	CCATGGACCTCC-----GAGCAGGTTGAAACCCCAACTAGCCCTCTTCATAA 168
Db	
382 GluLeuArgSerLeuProProAspMetValAlaGlyProArgLeuProAspThr----- 399	
QY 169	CATGACAAGCAGCGCCCATCTGATACCTAAAC-----CGACCAAGTCACAGCCCTCCA 222
Db	
400 -----PheLeuGlySerAlaThrProLeuHisPheProPro 411	
QY 223	ACTCAGCCTCTGCTGCCAGCCTCACACATCTCTTGTGGACTCAAACTCAACCCGAC 282
Db	
412 Ser---ProPheProSerSerThrGlyProGlyPro----- 422	
QY 283	TAAATCAACCAATCCCAAGTCTAAACTATCTGAAACTTTTAAAGTAAGCCAGTCCTTA 342
Db	
423 -----HisTyLeu 425	
QY 343	AACTAACTAGCCCAATGCCAATTATCTACCTAGCCAAACCTTAAGCTTTGCC 402
Db	
426 SerGlyProLeuProGlyThrTySerGlyProThrGlnLeuLe----- 441	
QY 403	AGTCMAAGTGCTCACTGAATCTCACTTGTGCTCAGTGAATAATCCAGAAAAGCATA 462
Db	
442 GlnProArgAlaPro-----GlyProHis-----Ala 450	
QY 463	TTTCCCACTGGCCACATCCCTCTTACAGCACCCA-----ACCCTGGCCTCT 510
Db	
451 MetProValAlaProGlyProAlaLeu---TyrProAlaProAlaTyThrProGluLeu 469	
QY 511	GCACCTCTGGTATCTCCGGGATGCCAACTCTGCAGTGCCATCAGCAACAAAGCCGACT 570
Db	
470 GlyLeuVal-----ProArgSerSerProGln 478	
QY 571	CGTCAATGCACCTCTCTCCCTTCCTG-----TCCCACTCTGAGGCTGAT 618
Db	
479 HisGlyValValSerSerProTyThrValGlyValGlyProAlaProVal----- 495	
QY 619	GGAAAGGCTCATTTGAAGTCCAACTTTTCCCACTTAACCAAGACGGGGTGAACCTC 678
Db	
496 -----AlaGlyLeuProSer-----AlaProProGlnPheSerGlyProGluLeu 511	
QY 679	CACACTGCCACCGTTCCTCGAGAGTGAGCACTAAATCTCTCAATCTAACCCCACTCA 738
Db	
512 AlaMetAlaValArgProAlaThrThrValAspSerIleGlnAla----- 527	
QY 739	CACCTCCACACTCAGGAATCACATCTAGATATATACCCAAATAAGCCCAATAAGGCA 798
Db	
528 -----ProIleProSer-----HisThrAla 534	
QY 799	GCCGACCTTAGTGTCTAACCTTATACCTTGTCTCTATGGGTGAGTCTGTTCTTGGC 858
Db	
535 ProArgPro-----AsnProThrProAla----- 542	
QY 859	GCGGCTCTCTCTGCTTCTCTCCCTTAGAGCTGACTGTGCTCAGCTGCCAGCTCTGACA 918
Db	
543 -----ProProProPro 546	
QY 919	TGTGCTGTCTCCCACTCTGACTCCCTCAAGCTGCAGTGGGACTGGAGACTGGCAGG 978
Db	

Db	547	CysPheProValProProProGlnProLeuProThrProTyThrTyThrProAlaGlyAla 566
QY	979	AAGCTAGGTTACAACTGGAAACACAGGAGGTCGACCTGCAGTCCCTAGGCTGGCCCT 1038
Db	567	Lys-----GlnProIle 570
QY	1039	CCCTCCATGTACACATATATATGTTGGCACACACAGTGGGCACACATGCAAGACT 1098
Db	571	ProAlaGlnHis---HisPheSerSerGlyIleProThrGlyPheProAlaProArg--- 588
QY	1099	CTCTCAGCTGACACACAGATCCATTCAGTATCTACTGATAGACACTCATCGCGTCCA 1158
Db	589	-----IleGlyProGln-----Pro 593
QY	1159	AGTCTCATCTCTCAACATACATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1215
Db	594	GlnProHisProGlnProHisProSerGlnAlaPheGlyProGlnProGlnProGlnPro 613
QY	1216	TTTCCCTCTCTCCATCCCTCTGCTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1275
Db	614	LeuProLeuGlnHisProHisPheProGlnAlaProGlyLeuLeuProProGln 633
QY	1276	CCCAGGTGGGACACAGACACTGAGGGCTGCCAGCTCTCCCGTCTGGGCCCGGGCC 1335
Db	634	-----SerProTyPro----- 637
QY	1336	CGCGTCATGCTTCTCGTCCATCTCTGCCACAGGAGCTCGTGGGGATGTTAGCGTGCCT 1395
Db	638	-----TyrAlaProGlnProGlyValLeuGlyGlnProProProPro 651
QY	1396	GTGCACGGTGTCTGGCACCTCCCTGTCAGTCCAGCTCTCAATCGCACAGGAGCCAGG 1455
Db	652	LeuHisThrGlnLeuTyPro-----GlyProAla 661
QY	1456	GCCTGGCCCTC-----CATCCAGAAAACCTATGACCTCACCTGCTACCTCGAGCACCA 1509
Db	662	GlnAspProLeuProAlaHisSerGlyAlaLeu---ProPheProSerProGlyProPro 680
QY	1510	ACTCCGAGCTTGGTGGGACCTATGATATCCAGGTAGGAATCTGGGAGTTGGGGA 1569
Db	681	GlnPro-----ProHisProProLeuAlaTyThrGlyProAlaPro 693
QY	1570	GGAGTGGAGGTGGGAAAGACAGTC-----CTAACGGGAGGGTCTCGTAAATGA 1623
Db	694	SerThrArgProMetGlyProGlnAlaAlaProLeuThrIleArgGlyProSer----- 711
QY	1624	TGGGTGAGGAGGGCTCTTTGGTCCCAACAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCT 1683
Db	712	-----SerAlaGlyGlnSerThrProSerProHisLeuValPro 724
QY	1684	TTCCCTCTTAGGTGGCCCGCCCTCTCCCATCTCCCGCCCGCCAGGACTAGGCATGTGGC 1743
Db	725	SerProAla-----ProSerProGlyPro----- 732
QY	1744	AGGCTCGCACCCCTTGGCCCATTTGCCACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1803
Db	733	-----GlyProValProProArgProProAla-AlaGluProProPr 746
QY	1804	CCCTGGGGCGGGAGTCT 1845
Db	746	o-----CysLeuArgArgGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 758
QY	1846	GGTGTCTCTTGGCGGGGGTGGGTGGGACAGAGGGGCCACCTCCCATGTGCTGC 1905
Db	758	uSerSerSerProGluSerGlnHisGlyGlyThrGlnSerProGlyGlyGlyGlnProLe 778
QY	1906	GTTCAG-----CTCGCTCTGCCCCCAGACCTGGGGCTCTGCTGCTGCTGCTGCTGCTG 1950
Db	778	uLeuGlnProThrIysValAspAlaAlaGluGlyArgArgProGlnAlaLeuArgLeu1 798
QY	1951	GACCCAGGGCT-----CCCTTCCTGCTG---CCTCTCCCATCTCTAGTGGCCT 1998
Db	798	eGluArgAspProTyThrGluHisProGluArgGlnLeuGlnGlnGluAl 818

Qy	1999	CCTAGGGGGT	CATGGGGAAAGGGGACTGTAGGAAACCCAGGACAGTAGTGCCAGGGGGTT	2058
Db	818	aPhe	::::: ArgGlyGlnLeuGlyAspValGlyAlaLeuAsp	830
Qy	2059	TAGGTGTGGAT	TGGAGTTATGCTGTAAAGATTGGGGGTGGTCCAGAGGTGTCACAGA	2118
Db	831	-ThrValTrpArgGluLeu	::::: GlnAspAlaGlnG	841
Qy	2119	CCCCAGGAGAGA	AGGAGGGTGGAGAGCCGAGG	2160
Db	841	uHisAspAlaArgGlyArgSerIleAlaIleAla-ArgCysTyrSerLeuLeuLysAsnArgH	861	
Qy	2161	ATGGGGAA	-----CCGGCCCTCTTCCGGTGTTCTCTTCACATCCAGACCTACT	2214
Db	861	IsGlnAspValMetProTyrAspSerAsnArgValValLeu	::::: ProGluLeuGlyLeuP	874
Qy	2215	CTGAGCCAGGAA	AGAAAGAGAGTGGCGGGAGCTGGCTCCAGCCCCAGGAT	2274
Db	875	-----ArgSerGlyLys	-----	878
Qy	2275	ACACGAGAAAT	TGTTGTCTGTGTGTGCAGGTGTGAACCTCCCTCGGCCCT	2334
Db	879	-----AspAspTyrIleAsnAlaSerCysValGluGlyLeuSerProTyrCysProProL	897	
Qy	2335	TG	-----CCTATCCAGGCTCTCCCTTGCCTTCTCCCTCTCTTCCAGT	2379
Db	897	euValAlaThrGlnAlaProLeuProGlyThrAlaAlaAspPheTrpLeuMet	::::: Met	914
Qy	2380	TATACATCTCCCTCAT	CTCTTCCCTGGGCCCCAGCGCTCCCCCGAGGTGGAAAGGG	2439
Db	914	-----	-----	914
Qy	2440	CTCTGCCCTCTTCCCTATACCA	NGTGTCTTCATAGCCTTCTCTGTCTCTACTCATGA	2499
Db	915	-----ValHisGluGlnLysValSerValIleValMet	925	
Qy	2500	GACTGCCTCCATTTCTCTCTCTGCAACCTGCTCCTATCAGCTGAACCTTCTTTCGGA	2559	
Db	926	-----LeuValSerGluAlaGluMetGluLysGlnL	936	
Qy	2560	GTGTTAGTAGTACCGCTCTCTCCCGCCCTCAGCTGGTGGCCCTGGGTGTGCAGCG	2619	
Db	936	ysValAlaArgTyr	-----PheProThrGluArgGly	946
Qy	2620	GCAATGGGGCTCTGTTCCAA	TGGCCACTCTCATCTCTCTGTTCTTGTGTGCAGAA	2679
Db	947	-----	-----G	947
Qy	2680	AACCTTGTCTTCCAC	-----TCCACTGCCTCTCTAGTTCCCGACCC	2720
Db	947	lnProMetValHisGlyAlaLeuSerLeuAlaLeuSerSerValArgSerThrGluThrH	967	
Qy	2721	-----TTTTTCTCTCTGGCTTTCCTTGC	CAAAATTTCTCAAGAGAGTGTCTAC	2769
Db	967	IsValGluArgValLeuSerLeuGlnPheArgAspGlnSerLeuLysArgSerLeuValH	987	
Qy	2770	ACCCTCTGCCTCCACTCTCTCTCCACCCTACTCACTTCTTAA	CCCCCTCGAATCTGGCTTC	2829
Db	987	Is	-----LeuHisPheProThrTrp	999
Qy	2830	CA	-----GGCCCGCAGCAATGTTCTCTCCAAGTTC	2874
Db	999	roAspSerProSerAsn-LeuLeuArgPheIleGlnGluValHisAlaHisTyrLeuHis	1018	
Qy	2875	GCCAAAGCCGACAGTGTGTTGAAGGCTCATCTCTCTGTGTGTGTTGTTTCGACCCACACT	2934	
Db	1019	GlnArgPro	-----LeuHisThrProIleIleValHisCysSerSerGly	1033
Qy	2935	GCTGAGCGCTGTCGCTTCTCGAACTCTCTTCTCTGTTGTGTCTGCACTCTCTCGGGCCAC	2994	
Db	1034	ValGlyArgThrGlyAlaPheAlaLeuLeuTyrAlaAla	-----	1046

[illegible]





```
Db 199 -----ArgGlySerArgMetGlyArgGlnHisGluAla---AlaAlaThrAlaAla 214
Qy 2503 AGTCTCATGATGAGGACAGGAGCGCTATGAGACAGCATGGTATAGGAGAGGGC 2444
Db 215 ThrAlaAlaThrAlaThrGlyGlyThrAlaGluGluAlaGlyAlaSerAlaProGluSer 234
Qy 2443 AGAGCCCTTTCCAAACCTCGGGGAGCGGCTGGGGCCCAAGGAAAGGATGAGGGAGATG 2384
Db 235 GlnAlaGlyGlyGlyProArgGlyArgAlaArgGlyProArgGlnGlnGlyArgArg 254
Qy 2383 TATACTGGGAAGAGG-----GAGNAGCAAGGGGAGAGCCCTGGG 2342
Db 255 HisGlyThrGlnArgArgArgGlyProProGlnAlaArgGluGluGlyProArgAspAla 274
Qy 2341 -----ATAGGCAAGGCG---CCAGGGGAGGTTCCACCGCTGACAGACAGAGAC 2294
Db 275 ThrThrIleGluGlyLeuGlyThrProSerGlyGluGlnArgAlaAspGln----- 291
Qy 2293 AAACTAATTCTCGGTGTATCCT-----GGGGCTGGAGCCAGCTCCCGCCGACCC 2243
Db 292 -----SerGlnAlaLeuProAlaLeuAlaGlyAlaAlaAlaHis--AlaHisA 308
Qy 2242 TCTCTCCCTTTCTTCCCTGGCTCCAGAGTAGGGTCTGGATGTGGAGGAGGACACGG 2183
Db 308 laileProGlyAlaGlyProAlaAlaAlaProValGlyArgGlyArgGlyGlyT 328
Qy 2182 GAAGAGGGGGCGGTTCCCATGCTCGGCTCGCTCCCAACCTCCTCTCTCTCTCT 2123
Db 328 rPArgGlyGlyArg----- 332
Qy 2122 GGGCTCTCTGAACACCTCTGGACCACCCCAATCTTACAGCATAACTCCATCCACAC 2063
Db 332 ----- 332
Qy 2062 CCTAAACCCCTGCCACTACTGCTGGTTCCTACAGTCCCTCCCATGACCCCCC 2003
Db 333 -----A 333
Qy 2002 TAGGAGGCCAGCTAGGATGGGAGAGCAGACGGAAGGAGGCCCTGGGTCCAGAGCAG 1943
Db 333 rgGlySerAlaGlyAlaGly---GlyGlyGlyArg-----GlyGlyArgGly- 348
Qy 1942 CAGGGCCCGGCTCTGGGGCAGAGCGAGCTGGAACGAGGATGGAGGTGGGGCCCC 1883
Db 349 ArgGly---ArgGlyGlyArgGly---GlyGlyGlyAlaGlyArgGlyGlyGlyAla-- 366
Qy 1882 TCTGTCCCAACCCACCCCGCGGAGAGACACACACACGCTGTAAACAGAGAGG 1823
Db 367 ----AlaGlyProArgGlyAlaSerSerProGlyAlaArgArgGlyGluGlnArgAr 385
Qy 1822 ACTTCCCGCG---CCCCAGGGGAGGCGGG-----CGGCT 1790
Db 385 gArgGlyArgGlyProProAlaAlaGlyAlaAlaGlnValSerAlaArgGlyArgAl 405
Qy 1789 GGGCTGGCAGCCAGTGGGGAATGGGCAAGGGGCTGGAGGGCTGCCACATGCCCTAG 1730
Db 405 aArgGlyGlnArgAlaGlyGluGluAlaGlnAspGlyLeuLeuProArgGlyArgAspAr 425
Qy 1729 TCTTGGGGCAGGGATGGGAAGTGGG-----GGGCCACTAGAG 1688
Db 425 gLeuProLeuArgProGlyArgAlaAlaAsnGlnArgAlaGluArgProGlyProArg-- 444
Qy 1687 GGAAGGCGAGGATAGACACAGAGGGGACTGGTGGGAGCCAAAGAGCCCTCCTCAC 1628
Db 445 -----GlyGly- 446
Qy 1627 CCCATCATTTACAGAACCTCCACGGGTAGGACTGTCTTTCCCAACTCCTCACTCCTC 1568
Db 447 -----HisGlyProValAsnAlaSerSerAlaProAspThrSe 459
Qy 1567 CCCAACTCCAGATTCTACGCTGGATCTACATAGGTCACGCCCAAGCTGGGAGTTG 1508
Db 459 rPro-----ProArgHisProArgArgtr 467
```

```
Qy 1507 GTGCTCCAGGTAGCGCGGTGAGGTTCATAGTTTTCGTGATGAG----- 1465
Db 467 pValSerGlnGlnArgGlnArg-----LeuTrpArgGlnPheArgValGlyGlyG 484
Qy 1464 -----GGGCCAGGCCCTGGTCCCT-----GTCGGATTGAG 1433
Db 484 yPheProProProProSerArgProProAlaValLeuLeuProLeuLeuArgLeuAl 504
Qy 1432 A---GCTGGCACTCAGGGAGGTGCCAGACACCGTCACAGCAGCGTAACTCCCCCA 1376
Db 504 aCysAlaGlyAspProGlyAlaThrArgProGlyProArgArgProAlaArgArgProAr 524
Qy 1375 C----- 1375
Db 524 gGlyGluLeuIleProArgArgProAspProAlaAlaProSerGluGluGlyLeuArgMe 544
Qy 1374 -GAGTCCCTGTG-----GGCAGATGGAGAGCATGAGCGCGGCCGCCACCA 1322
Db 544 tGluSerSerValAspAspGlyAlaThrAlaThrThrAlaAspAlaAlaSerGlyGluAl 564
Qy 1321 CGGGGAAGCAGCTGCAGCCCTCAGGTGTCTGTCCCACTTGGCTGGGTGGGGGTG 1262
Db 564 aProGluAlaGlyProSerProSerHis-Ser---ProThrMetCysGlnThrGlyGlyP 583
Qy 1261 AGGTGGGACACCA 1248
Db 583 roGlyProProPro 587

RESULT 8
US-09-738-973-187
; Sequence 187, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Segrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 187
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-738-973-187

Alignment Scores:
Pred. No.: 1.98e-06 Length: 595
Score: 239.50 Matches: 184
Percent Similarity: 33.29% Conservative: 59
Best Local Similarity: 25.21% Mismatches: 240
Query Match: 2.57% Indels: 247
DB: 10 Gaps: 39

US-09-931-704-3 (1-5087) x US-09-738-973-187 (1-595)
Qy 3130 AAAGGAGGAGGA-----GGAGACGACAGAGCGCTCAGGAGGACGTGCTGAT 3083
Db 8 GluGlyGlyGlyGluTrpGlyProGlyValProArgGluArgGluSerAlaGlyGlu 27
```







QY 2408 CCCAGCGCTCCCGGAGGCTGGAAAGGCTCTGCCCTCTTCCCTATACCATGCTGC 2468  
Db 1468 hrThrProSerProPro-----ThrThrThrProSerPro----- 1479  
QY 2469 TTCATAGCTTCCTCCCTCTACTCANGAGCTGCCTCATTTCTTCTTCTGCAACC 2528  
Db 1480 -----ProThrThrThrProSerProProThrThrThrThr-----P 1494  
QY 2529 CTGCTCTATCAGCTGAACCTTCTTTCCGGAGTGTAGTGAGTACCCGCTCTCCCCAGC 2588  
Db 1494 roProProThrThr-----:ProProSerP 1502  
QY 2589 CCCTCAGCTGGTGGGCTGGGTGTGTCTCAGCGCAAAATGGGGCTCTGTTCCAAATGGGCCA 2648  
Db 1502 ro----- 1502  
QY 2649 CTCTCATCTCTCTTGTTCCTTGTGCGAAGAAACCTTTGCTTCACTCCACTGCCCTCTCT 2708  
Db 1503 -----PrometT 1505  
QY 2709 AGTTCCCGACCTTTTCTCTCTCTGCTTCCCTGCCAAATTTCTCCAAGGAGTGTCTA 2768  
Db 1505 hrThrProile-----T 1509  
QY 2769 CACCTCTGCTCTCACTTCTCTCT--CCACCACCTCACTTCTTAAACCCCTGCAATCTGG 2825  
Db 1509 hrProAlaSerThrThrLeuProProThr-----ThrThrPro----- 1523  
QY 2826 CTTCAGGCCCCAGCAATGGTTCCTCCAGGTCCTCAGGACCTCTTGGCAAGCCGGA 2885  
Db 1524 -----SerProProThrThrThrThrThrThrProProProT 1536  
QY 2886 CAGTGTTTTGAAGGCTCATCTCTCTCTGCTGTCTGTTTTCAGCCACACTGCTGAGCGCTG 2945  
Db 1536 hr----- 1536  
QY 2946 CTGCCTTCTCGAATCTCTTCTTGGTCTCTGACTCTCTGGGCGCACCTTCTACCTCT 3005  
Db 1537 -----Thr-ThrProSerProProThrThrThr 1545  
QY 3006 CAGCTCTCTCAGGCT 3065  
Db 1546 ProilerProProThrSerThrThrThrLeuProProThrThrThrProSerPro--- 1564  
QY 3066 GTTTCGCCACCCAGCAATCAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3125  
Db 1565 -----ProProProThrThrThrThrThrThrThrThrThrThrThrThr 1582  
QY 3126 CTTTTTCTACGCTCTCCA-----TTGGAGAGCTCACACCGGCACTGCTTCAAC 3176  
Db 1583 ThrThrThr--ProSerProProThrThrThrThrThrThrThrThrThrThrThr 1602  
QY 3177 TGTACCTGCATACAAATGATATCTTATTGAAAACTCAGGAGGCGCATGAAACAAGA 3236  
Db 1602 roSerPro-----ProThrThrThrThrThrThrThrThrThrThrThr 1617  
QY 3237 AGCCTAGC--ATGGAGACAGGCGCAGTGTCTCAGGGACACAAAAAATAGAACTTTGGGA 3293  
Db 1617 hrProSerProProThrThrThrThrThrThrThrThrThrThrThrThrThrLeu--- 1635  
QY 3294 GCAGTATCTCTCTGTGTGTGAGCCAGCGGCTCTGCGCTCTCTCTCTCTCTCTCTCTCTCT 3353  
Db 1636 -----ProProThrThrThrThrThrThrThrThrThrThrThr 1642  
QY 3354 CTTTTTCAGCTGAATCTAGTGGGCGCCCTTTTCAAGAGGACACTTCAACCTCCCTCC 3413  
Db 1643 ProProProThrThrThrThrThrPro-----ProThrThrThrPro 1657  
QY 3414 GCCTGGGCGCAGACTCTGCCAGGCGGCACTGTTGACTTTGGAGTGTGGCGAAGCTCA 3473  
Db 1658 SerProProThrThrThrThrProSerProProThrThrThrThrThrThrThrPro 1677  
QY 3474 ATGACAAACTGGGCTGACCCAGAACTACGAGGCGCTTACAGCCACCTTCTGTGTACTTGC 3533

Db 1678 ThrThrThrProSerSerProThrThrThrThrThrThrThrThrThrThrThrMetTh 1697  
QY 3534 GTGGCTCTCAACCGTCAGGCTGCCACTGCTGAGCTGGCGGAGCTGGCCCTCTCTCA 3593  
Db 1697 rThrProSerProThr-----ThrThrProSerSerProThrThr 1710  
QY 3594 CCAGCTCCAGGCGCTGCTGGGAGCAATTCGGGGCTCATGGCAGCTCTGGGCTACCCAC 3653  
Db 1710 rThrThrThrThrPro-----SerSerThrThrThr 1720  
QY 3654 TGCCCCAGCCCTGCTGGGACT-----GAACCCACTTGGACTCTGGCCCTG 3701  
Db 1720 r-ProSerProProThrThrThrMetThrThrProSerProThrThrThrProSerPro- 1739  
QY 3702 CCCAGTGAAGTCTCTCCAGAGATGGAGCACTTCTGCTGCTGAAGGAGTGCAGACT 3761  
Db 1740 ProThrThrThrMetThr-----ThrLeuProPro 1749  
QY 3762 GGCTGTGGGCTCGGCCAAGACTTCAACCGG-CTCAAGAAAGATGAGCTCCAGCA 3820  
Db 1750 ThrThrThrSerSerProLeuThrThrThrProLeuProProSerThrThrProPro--- 1768  
QY 3821 GCTGAGTCACTGACCTGGGCTCATGGCTCTGACTTCTGACTTCTCTCTCTCTCTCG 3880  
Db 1769 -----ThrPheSer----- 1771  
QY 3881 CTCCTCTTCAACCTCTGCTCCCTTGTGAGAGCAGCCCTGTATGCAACACTGTGT 3940  
Db 1772 ---ProPheSerThrThrThrPro-----ThrThrProCysValPro----- 1784  
QY 3941 GAGCCAGGAGACAGAAAGCTGTGAGCCTCTGGCCCTTCTCTGACCGGCTGGCGGTGTGAT 4000  
Db 1785 -----LeuCysAsnTrpThrGlyTrp---LeuAsp 1793  
QY 4001 GGGATCAGCCCTGTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 4039  
Db 1794 SerGlyLysProAsnPheHisLysProGlyGlyAspThrGluLeuIleGlyAspValCys 1813  
QY 4040 ---CCGAGCTCGGGAGGA----- 4054  
Db 1814 GlyProGlyTPAlaAlaAsnIleSerCysArgAlaThrMetTyrProAspValProIle 1833  
QY 4055 GGTACAGTAGGCGCTCTGCT 4114  
Db 1834 GlyGlnLeuGlyGlnThrValValCysAspValSerValGlyLeuIleCysLysAsnGlu 1853  
QY 4115 GTGGTTCAGGTTGGTGGAGCGCTCATGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4156  
Db 1854 AspGlnLysProGlyValIleProMetAlaPheCysLeuAsnTyrGluIleAsnVal 1873  
QY 4157 -----TTGCTACCTTGGCCAGTG-----CCC 4180  
Db 1874 GlnCysCysGluCysValThrGlnProThrThrMetThrThrThrThrThrGluAsnPro 1893  
QY 4181 ACCCAGCCCTCAGGTGGCAGCATCTGGAGGCGAGGGGTGAGGGGCCACACACACATG 4240  
Db 1894 ThrProProThrThrThrProIleThrThrThrThrThrValThrProProPro--- 1911  
QY 4241 CCTTTCTGGGTGAAGCCCTTTGGCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4300  
Db 1912 -----ThrProThrGlyThrGlnThrProThr 1920  
QY 4301 TCCCCCAATCACTATATACATCCAAATTCAGGAAACAAACATGGTGGCAATTTCTACAAA 4360  
Db 1921 ThrThr-----ProIleThrThrThrThrValThr----- 1931  
QY 4361 AAGAGATGAGATTAAACAGTGCAGGGTGGGGTCTGCTGCTTGGAGGTGCCCTATAAACAGA 4420  
Db 1931 ----- 1931  
QY 4421 AGAGAAATATCTGAAGACAGAGGGGAGGAGACAGACAGACAGACAGACAGAGAGTCTCCA 4480  
Db 1931 ----- 1931



QY 1856 TCGCGGGGGGGTGGGTGGGACAGAGGGGCCCCACCTCCCATGCGCTCCGTTCCAGCT 1914  
Db 1332 -----AlaProGluAspIleGluCysArgSerVal 1341  
QY 1915 CGCCTCTGCCCCAGACAGCTGGGGCCCTGCTCTCGAGCCAGGGCCCTCCCTCCGCTCT 1974  
Db 1342 Lys-----AspProHisLeuSerLeuGluGln 1350  
QY 1975 GCCTCTCCCATCTAGCTGGGCTCTAGGGGGTTCATGGGGGAAGGGGAC----- 2025  
Db 1351 -----HisGlyGlnLysValGlnCysAspValSerVal 1361  
QY 2026 -----TGATAGGAACCCAGGCGATGATGCGAGGGGTTAGGTGTGGATGGAGGT 2076  
Db 1362 GlyPheIleCysLysAsnGluAspGlnPheGlyAsnGlyProPheGlyLeuCysTyrAsp 1381  
QY 2077 TATGCTGTAAGGAT-----TGGGGGTGCTCAGAGGTCTTTCAGAGAGCCAGGAG 2127  
Db 1382 TyrLysIleArgValAsnCysCysIleP-----ProMetAspLysCysIleThrThrPro----- 1399  
QY 2128 AGAAGGAAGGAGGTGGAGGAGCCAGGCACCATGGGGAACCGGCCCTCTTCCCGTG 2187  
Db 1400 -----SerProProThrThrThrProSerProProProThrThrT 1414  
QY 2188 TTCTCTTCCCATCCAGACCCCTACTCTGAGCCAGGGAAGAAAGGAAGAGGTGG 2247  
Db 1414 hrThrLeuPro----- 1417  
QY 2248 CGGGGAGCTGGCTCCAGCCCCAGGATACACCCAGGAAATAGTTGTCTCTGTCTGT 2307  
Db 1418 -----ProThrThrThrProSerProProThrThr-----T 1428  
QY 2308 CAGCGTGTGAACCTCCCTCCCTGGGCCCTTGCTATCCAGG-----C 2348  
Db 1428 hrThrThrThrProProProThrThrThrProSerProProIleThrThrThrThrP 1448  
QY 2349 CTCTCCCTGTGTCCTCTCTTCCAGTTATACATCTCCCTCATCTCCCTTCCCTGGG 2408  
Db 1448 roLeuProThrThrThrProSerProProIleSerThrThrThrThrProProThrT 1468  
QY 2409 CCCCAGCGCTCCCGAGGGTGGAAAGGCTCTGCCCTCTCCCTATACCATGCTGTC 2468  
Db 1468 hrThrProSerProPro-----ThrThrThrProSerPro----- 1479  
QY 2469 TTCCATAGCTTCTCTCTCTCTACTCATGAGAGTGCCTCATTTCTCTCTCTGCAACC 2528  
Db 1480 -----ProThrThrThrProSerProProThrThrThrThrThr-----P 1494  
QY 2529 CTGCTCTATCAGCTGAACCTCTCTTTCGGAGGTGTAGTAGTACCCGCTCTCTCCCGAGC 2588  
Db 1494 roProProThrThr-----ThrProSerP 1502  
QY 2589 CCCTCAGCTGGTGGGCTGGGTGTGTGACGGGCAATGGGGCTCTGTTCCAATGGGCCA 2648  
Db 1502 ro----- 1502  
QY 2649 CTCTCATCTCTCTGTGTTCTCTCTGCTTTCCTGCTTTCCTGCAAAATTTCTCAAGGAGTGGTCTA 2768  
Db 1503 -----ProMetT 1505  
QY 2709 AGTTCCCGACCTTTTCTCTCTCTGCTTTCCTGCTTTCCTGCAAAATTTCTCAAGGAGTGGTCTA 2768  
Db 1505 hrThrProIle-----T 1509  
QY 2769 CACCTCTGCTTCCACTTCTCT-----CCACCGACTCACTTCTTAAACCCCTGCAATCTGG 2825  
Db 1509 hrProProAlaSerThrThrLeuProProThr-----ThrThrPro----- 1523  
QY 2826 CTTCAGGGCCCCAGCAATGGTCTCTCCAAAGTCTGTCAGGACCTCTCTTCCCAAGCCCGA 2885  
Db 1524 -----SerProProThrThrThrThrThrThrThrThrThrThrThrPro 1536  
QY 2886 CAGTGTTTGAAGGCTCATTTCTCTCTGCTGTGTTTTCAGGACCACTGCTGAGGGCTG 2945

Db 1536 hr----- 1536  
QY 2946 CTGCTTCTCGAACTCCTCTTCTTGTGCTCTGCACTCTCTGGCCGACCTTCTACCTCT 3005  
Db 1537 -----ThrThrProSerProProThrThrThr 1545  
QY 3006 CCAGCTCTCTCAGGCTCCTCTTCTCTCTCTGCTCTCCGCCACAGCGGGGCACTCTCCCAAG 3065  
Db 1546 ProIleThrProProThrSerThrThrThrLeuProProThrThrThrProProPro----- 1564  
QY 3066 GTTTGGCCACCCAGCAATCAGACGCTCTCTCCTGAGCGTCTTGTGCGTCTCTCTCTCTCT 3125  
Db 1565 -----ProProThrThrThrThrThrThrProProProProThrThrProSerProThr 1582  
QY 3126 CCTTTTCTAGCCCTCTCCA-----TTGGAGAGCTCACCGGCGCACTGCTTCAAC 3176  
Db 1583 ThrThrThr-----ProSerProProProIleThrThrThrThrProProThrThrThrP 1602  
QY 3177 TGTCACTTGCATACAAATGATATCCTTATTGAAAAAATCTCAGGAGGCCATGAACAAAGA 3236  
Db 1602 roSerPro-----ProThrThrThrThrThrThrProProProProThrT 1617  
QY 3237 AGCTAGC-----ATGGAGACAGGGCCAGTGTCTAGGGGACACAAAAAATAGAACTTTGGGA 3293  
Db 1617 hrProSerProProThrThrThrProIleThrProProThrSerThr-----ThrThrLeu----- 1635  
QY 3294 GCAGGTATCTCTTGGTGGTGAGCCAGCGGCTCTGCCCTCTCTCTCTCTCTCTCTCTCT 3353  
Db 1636 -----ProProThrThrThrThrProSer 1642  
QY 3354 CCTTTTCAAGCTGAACCTACCTGGGCCCTCTTCAAGAGCCAGACTTCAACCTCTCCCC 3413  
Db 1643 ProProProThrThrThrThrThrPro-----ProThrThrThrPro 1657  
QY 3414 GCCTGGGGGAGAGACTCTGCCAGGCGCACTGTGTGACTTGGAGGTGTGGGAGGCTCA 3473  
Db 1658 SerProProThrThrThrThrProSerProProIleThrThrThrThrProProPro 1677  
QY 3474 ATGACAACTGCGGCTGACCCAGAACTAGAGGCTACAGGCGCTACGCCACTTCTGTGTACTTGC 3533  
Db 1678 ThrThrThrProSerSerProIleThrThrThrThrThrThrThrThrThrMetTh 1697  
QY 3534 GTGGCTCAACCGTCAGGCTGCCACTGTGAGTGTGCGGCGAGCTGGGCCACTTCTGCA 3593  
Db 1697 rThrProSerProThr-----ThrThrProSerSerProIleTh 1710  
QY 3594 CCAGCTCCAGGCGCTGCTGGGCGAGCAFTGCGGGCGTCATGGCAGCTCTGGGCTACCCAC 3653  
Db 1710 rThrThrThrThrPro-----SerSerThrThrThr 1720  
QY 3654 TGCCCCAGCGCTCGCTGGGACT-----GAACCCACTTGGACTCTGGGCCCTG 3701  
Db 1720 r-ProSerProProProThrThrMetThrThrProSerProThrThrProSerPro- 1739  
QY 3702 CCCACAGTACTTCTCTCCAGAAGATGGACGACTTCTGGTGTCTGAAGGAGTGCAGACCT 3761  
Db 1740 ProThrThrThrMetThr-----ThrLeuProPro 1749  
QY 3762 GGCTGTGGCGCTCGGCCAAGGACTTCAACCGG-CTCAAGAGAAGATGAGGCTCCAGCA 3820  
Db 1750 ThrThrThrSerSerProLeuThrThrThrProLeuProProSerIleThrProPro----- 1768  
QY 3821 GCTGCAGTACCCTGCACCTGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCT 3880  
Db 1769 -----ThrPheSer----- 1771  
QY 3881 CTCCCTCTCAAACTGCTCTCCACTTTGTGAGAGCCAGCCCTGTATGATGCCAACCTGTT 3940  
Db 1772 -----ProPheSerThrThrThrPro-----ThrThrProCysValPro----- 1784  
QY 3941 GAGCCAGGAGACAGAGCTGTGAGCCTCTGGGCCCTCTTCTCTGAGCCGCTGGGCGTGTGAT 4000  
:::|||||



```
QY 1474 GAAACCTA-----TGACCTCAC 1491
|||:||||
Db 1176 GluAspLeuLysCysValThrAlaAspLysCysGlyCysTyrValGluAspThrHis 1195
QY 1492 CGCTACCTGGAGCAACCACTCGCAGCTGGCTGGGACCTAGTGTAGTATCCAGCGTAG 1551
|||||:|||||
Db 1196 TyrProGlyAlaSerValProThr-----GluGluThrCysLysSerCysValCys 1213
QY 1552 GAATCTGGAGTGGGAGGAGTGGAGGATGGGGAAGACAGCTCTTA-----ACCGTG 1605
|||:|||||
Db 1214 ThrAsnSerSerGlnValValCysArgProGluGluGlyLysIleLeuAsnGlnThrGln 1233
QY 1606 GAGGTTCTGGTAATGATGG-----GGTGAGAGGGGCTCTTTGGCTCCACAG 1656
|||:|||||
Db 1234 AspGlyAlaPheCysTyrTrpGluIleCysGlyProAsnGlyThrValGluLysHisPhe 1253
QY 1657 TCCCCCTGTCTGTCTATCTCTGCTGCC-----TTCCCTCTTAGGTGCGCC 1701
|||:|||||
Db 1254 AsnIleCysSerIleThrArgProSerThrLeuThrThrPheThrThrIleThrLeu 1273
QY 1702 CCCCACTTCCCATCTCCCTGGCCCCAGAGTAGCATGTGGCGAGGCT-----CGCACC 1755
|||||
Db 1274 ProThrThrProThrSerPheThrThrThrThrThrThrThrProThrSerThr 1293
QY 1756 CGCCTTGGCCCATGTGCC-----CACTGGCTGCCAGCCCGCCAGCCG 1794
|||||
Db 1294 ValLeuSerThrThrProLysLeuCysCysLeuTrpSerAspTrpIleAsnGluAspHis 1313
QY 1795 CCGCTCCCTCCCTGGGGCGGGAGTCTCTCTTTACCGTGTGTGTGTCTCT 1854
|||:|||||
Db 1314 ProSerSerGlySerAspAspGlyAspArgGluProPheAspGlyValCysGly----- 1331
QY 1855 TCGCGGGCGGGGTGGGTGGGACAGAGGGGCCACCTCCCATGCGCTCGCTCCAGCT 1914
|||
Db 1332 -----AlaProGluAspIleGluCysArgSerVal 1341
QY 1915 CGCCTCTGCCCCAGACCTGGGGCCCTGTCTCTGGACCCAGGGGCTCCCTTCOGTCT 1974
|||
Db 1342 Lys-----AspProHisLeuSerLeuGluGln 1350
QY 1975 GCCTCTCCATCTAGCTGGGCTCTAGGGGGTCTAGGGGGTCTAGGGGAAGGGAC----- 2025
|||
Db 1351 -----HisGlyGlnLysValGlnCysAspValSerVal 1361
QY 2026 -----TGTAGGAACCCAGGAGTAGTGGCAGGGGTGGTGGATGGAGGT 2076
|||:|||||
Db 1362 GlyPheIleCysLysAsnGluAspGlnPheGlyAsnGlyProPheGlyLeuCysTyrAsp 1381
QY 2077 TATGCTGTAAAGATT-----TGGGGTGTCTCAGAGGTGTTCAGAGAGCCAGGAG 2127
|||:|||||
Db 1382 TyrLysIleArgValAsnCysCysTrp--ProMetAspLysCysIleThrThrPro---- 1399
QY 2128 AGAAGGAGGAGGTGGAGGAGCCAGGACCACCTGGGNAACCGGCCCTCTCCCGTG 2187
|||
Db 1400 -----SerProProThrThrThrThrProSerProProProThrThrThr 1414
QY 2188 TTCCTCTTCCACATCCAGACCTACTCTGGAGCCAGGGAAGAAAGGAAGAGGTGG 2247
|||||
Db 1414 hrThrLeuPro----- 1417
QY 2248 CGGGGAGCTGGCTCCAGCCCGCAGGATACACCGAGAAATAGTTGTCTGTCTGTGTGT 2307
|||:|||||
Db 1418 --ProThrThrThrProProProProThr-----T 1428
QY 2308 CAGCGTGTGAACCTCCCTGGCCCTTGGCCCTATCCAGG-----C 2348
|||
Db 1428 hrThrThrThrProProProProThrThrThrProSerProIleThrThrThrThrP 1448
QY 2349 CTCTCCCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2408
|||||
Db 1448 roLeuProThrThrThrProSerProProIleSerThrThrThrThrProProProThr 1468
QY 2409 CCCCAGCGCTCCCGAGGGTGGAAAGGGCTCTGCCCTCTTCCCTATATACCATGTCTC 2468
|||
```





QY 831 CTCTCTAGGTGAG-----TCTGTCTTGGCGCGCGCTCTCTCTGCT 875  
Db |||||:|||||  
99 LeuProIleGlyGluLeuAspSerAlaHisSerIleVal-----ThrLysThrProAsp 116  
QY |||||:|||||  
876 TCCTCCCTTAGAGTGAAGTGTGTCAGCTCCAGCTCTGACATGTGTCTCCACCC 935  
Db |||||:|||||  
117 LeuSerLeuAlaGlu-----GluThrGlyLeuSerAspThrGluCysAlaAspPro 134  
QY |||||:|||||  
936 -----TCTACTCCCTCAAGTGCAGTGGGACTGGAAGAC 971  
Db |||||:|||||  
135 PheGluValIleHisProGlyAlaSerArgProAlaGluAlaGlyValGlyIleGlnAla 154  
QY |||||:|||||  
972 TGGCAGGAAGCTAGGTTACAACTGGAACACAGCAGGAGTCGACCTGCAGTCCCTAGGCTG 1031  
Db |||||:|||||  
155 ThrAlaGluAspLeuThrAlaAlaGlyAlaGln---AspGlyGlnArgThrSerSer 173  
QY |||||:|||||  
1032 GCCCGTCCCTCCATGTATACACATATATACATGTTGGGCACACACAGTGGGCACATGCC 1091  
Db |||||:|||||  
174 GlySerAspThrAspIleTyrSerPhe-----HisSerAlaThrGluGln 188  
QY |||||:|||||  
1092 AAAGACTCTCTCAGCTGCACACAG-----ATCCATTCTCAA----- 1128  
Db |||||:|||||  
189 GluAspLeuLeuSer-AspIleGlnAlaIleArgLeuGlnGlnGlnGlnGlnGln 208  
QY |||||:|||||  
1129 -----GTATCTACTGATGACACTCATGCGTGCCA-----AGTCCTCATCC 1169  
Db |||||:|||||  
208 sLeuLeuGlnAspSerGluGluProAlaAlaProThrAlaIleSerProGlnPr 228  
QY |||||:|||||  
1170 T-----CA 1172  
Db |||||:|||||  
228 oGlyAlaPheLeuGlyLeuAspGlnPheLeuLeuGlyProArgSerGluAlaGluLysAs 248  
QY |||||:|||||  
1173 AACATACACATGCTCTCTCTCCGTCTTGCAGGAGTGTTCCTCTCCATCC 1232  
Db |||||:|||||  
248 pThrValGlnAlaLeuProValArgProAspLeuPro-GluThrThrLysSerLeuValP 268  
QY |||||:|||||  
1233 CC-----TCTGCTCCCATCTGTGTGCCACCTCACCCTCACCCTCACCCTCACCCTCACCCT 1277  
Db |||||:|||||  
268 roGluHisProProSerSerGlySerHisLeuThrSerGlu-----ThrP 283  
QY |||||:|||||  
1278 CAGGTGGGGACAGACACTGAGGGCTG-----CCAGCTGCTTCCCG----- 1321  
Db |||||:|||||  
283 roGlyTyrAlaThrAlaProSerAlaValThrAspSerLeuSerSerProAlaPheThrP 303  
QY |||||:|||||  
1322 -----TGTGGCGCGCGCGCTCATGCTTCTCGTCCATCTGCCACACAGGGACT 1373  
Db |||||:|||||  
303 heProGluAlaGlyProGlyGluGlyAlaAlaGlyValProValAlaGlyThrGlyAspT 323  
QY |||||:|||||  
1374 CG----- 1375  
Db |||||:|||||  
323 hrAspGluGluCysGluGluAspAlaPheGluAspAlaProArgGlySerProGlyGluG 343  
QY |||||:|||||  
1376 ---TGG-----GGCA 1382  
Db |||||:|||||  
343 luTrpValProGluMetGluGluAlaSerGlnArgLeuGluLysGluProGluGluGlyM 363  
QY |||||:|||||  
1383 TGTGTAGGTGCTGTGACAGGTGCTTGGCACCTCTCCAGTCCCTCAGTGCCAGCTCAATCGCA 1442  
Db |||||:|||||  
363 eArgGluSerIleThrSerAlaValAlaSerLeuPro----- 375  
QY |||||:|||||  
1443 CAGGGACCCAGGGCTGCGCCCTCCATCCAGAAACCTATGACCTCACCCTGCTGAG 1502  
Db |||||:|||||  
376 --GlySerProAlaProSerProArgCysPheLysProTyrProLeu----- 390  
QY |||||:|||||  
1503 AGCAACCACTCCGACGTTGGTGGGACCTATGTAGTAGTATCCAGCGTAGAATCTGGAG 1562  
Db |||||:|||||  
391 -----ileThrProCysTyrIleLysThrThrThrArgGlnLeuSerS 405  
QY |||||:|||||  
1563 TTGGGAGGAGTGAGGAGTTGGGGAAGACAGTCCTAACCGTGGAGGTTCTGTTAAATG 1622  
Db |||||:|||||  
405 erProAsnHis----- 408  
QY |||||:|||||  
1623 ATGGGGTGAGGAGGGGCTCTTTGGCTCCACCAGTCCCGCTGCTGTGTTATCTCTCTGCC 1682

Db |||||:|||||  
409 -----SerProSerGlnSerProAsnGlnSer-----P 418  
QY |||||:|||||  
1683 CTTCCCTCTTAGTGGCCCGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1742  
Db |||||:|||||  
418 roArgIleLysLysArgProAspProSerValSerProSerProArgThrAlaLeuAlaAs 438  
QY |||||:|||||  
1743 CAGGCTCGCACCGCGCTTGGCCCATTCGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1780  
Db |||||:|||||  
438 erAlaAlaAlaProAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 458  
QY |||||:|||||  
1781 -----TGCAGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 1826  
Db |||||:|||||  
458 rgSerAlaAspTrpThrGluGluLeuGlyValArgThrProGlyAlaGlyGlySer----- 476  
QY |||||:|||||  
1827 TCTGTTTACACCGTGTGTGTCTTTCGCGCGGGGGTGGTGGGACACAGAGGG 1886  
Db |||||:|||||  
477 -----ValHisLeuLeuGlyArgGlyAlaThrAlaAspSerGlyG 491  
QY |||||:|||||  
1887 CCCACCTCCATGCTGCTTCAGCTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1930  
Db |||||:|||||  
491 lyGlySerProValLeuAlaAlaLysAlaProGlyAlaProAlaThrAlaAspGlyPheG 511  
QY |||||:|||||  
1931 -----CCTGGGCGCTGCTGCTC-----TGCACCCAGGGC 1961  
Db |||||:|||||  
511 lnAsnValPheThrGlyArgThrLeuLeuGlyLysLeuPheSerGlnGlnGluAsnGlyP 531  
QY |||||:|||||  
1962 CTTCC----- 1966  
Db |||||:|||||  
531 roProGluGluAlaGluLysPheCysSerArgIleAlaMetGlyLeuLeuLeuProP 551  
QY |||||:|||||  
1967 -----TTCGCTGCTGCTCTCCCTCCCTAGCTGGGCTCTCCCTAGGGGGTTCATGGG 2015  
Db |||||:|||||  
551 heSerAspCysPheArgGluProCysAsnGln----- 561  
QY |||||:|||||  
2016 GGAAGGGACTGTAGGGAACCCAGCAGTGTAGGAGGGGGTGTAGGTGTGTGTGTGTGTGTGTGT 2075  
Db |||||:|||||  
562 -----AsnAlaGlySerSerAlaProPheAsp-----GlnA 573  
QY |||||:|||||  
2076 TTATCTCTAAGATTTCGGGTGTCCAGAGTGTTCAGAGAGCCAGCAGAGAGAGAA 2135  
Db |||||:|||||  
573 spGlnLeu-TyrThrTrpAlaValSerGlnProThrHisSerMetAspTyrSerGlu 592  
QY |||||:|||||  
2136 GGAGGTTGGAGGAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2195  
Db |||||:|||||  
593 Gly-----GlnPheProArgArgGluProSerMetTyrProSerSerLysLeuProGlu 610  
QY |||||:|||||  
2196 CCACATCCAGACCTCTCTGAGCCAGGGAAGAAAGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 2255  
Db |||||:|||||  
611 GluGluProSerProLysAspValAspThrGluProLysSerSerIleLeuGluSerPro 630  
QY |||||:|||||  
2256 CTGGCTCCAGCCCGCCAGGATACACAGGAAATT-----AGTTGTCT 2297  
Db |||||:|||||  
631 LysLysCysSerAsnGlyValGlnGlnGluValPheAspValLysSerGluGlnAla 650  
QY |||||:|||||  
2298 CTGTCTTGTACAGGTGTGAACCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2357  
Db |||||:|||||  
651 ThrValIleGlnGlnLeuGluGln-----ThrIleGluAspLeuArgThr 665  
QY |||||:|||||  
2358 TGCTTCT 2417  
Db |||||:|||||  
666 LysIleAlaGluLeuGluLysGlnTyrProAlaLeu---AspLeuGluGlyProArgGly 684  
QY |||||:|||||  
2418 CTTCCCGGAGGTGTGAAGGGCTCTGCTCTCT-----TCCCTATACCATGCT 2465  
Db |||||:|||||  
685 Leu---SerGlyLeuGluAsnGlyLeuThrAlaSerAlaAspValSerLeuAspAlaLeu 703  
QY |||||:|||||  
2466 GTCTTCCATGAGCT 2525  
Db |||||:|||||  
704 ValLeuHisGly-----LysValAlaGlnProProArgThrLeuGluAla 718  
QY |||||:|||||  
2526 ACCTGCTCTCTATCAGCTGAACCT 2576



```

; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 1806

```

Alignment Scores:	
Pred. No.:	4.77e-06
Score:	233.00
Length:	1806
Matches:	313
Conservative:	90
Mismatches:	477
Indels:	555
Gaps:	73
DB:	10

US-09-931-704-3 (1-5087) x US-09-919-497-56 (1-1806)

944	CCCTCAAGCTGCAGTGGGACGTGGAAAGATGCGCAGGAAGCTAGGTGTAACAATGGACACACAG	1003	                   	945	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1004	GCAGGTGCGACCTGCGAGTCCCTAGGCTGGC	1005	                   	946	GlyAspArgGlyProProGlyArgProGlyLeuProGlyAlaAspGlyLeuProGlyPro	1006	                   	947	GCAGTCCCT	1007	CCGCTCCCT	1008	                   	948	GlyAspArgGlyProProGlyArgProGlyLeuProGlyAlaAspGlyLeuProGlyPro	1009	                   	949	ProAlaGlyIleMetGlyProProGlyLeuGlnGlyProThrGlyProProGlyAspPro	1010	                   	950	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1011	GCAGTCCCT	1012	CCGCTCCCT	1013	                   	951	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1014	GCAGTCCCT	1015	CCGCTCCCT	1016	                   	952	GlyAspArgGlyProProGlyArgProGlyLeuProGlyAlaAspGlyLeuProGlyPro	1017	                   	953	GCAGTCCCT	1018	CCGCTCCCT	1019	                   	954	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1020	GCAGTCCCT	1021	CCGCTCCCT	1022	                   	955	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1023	GCAGTCCCT	1024	CCGCTCCCT	1025	                   	956	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1026	GCAGTCCCT	1027	CCGCTCCCT	1028	                   	957	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1029	GCAGTCCCT	1030	CCGCTCCCT	1031	                   	958	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1032	GCAGTCCCT	1033	CCGCTCCCT	1034	                   	959	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1035	GCAGTCCCT	1036	CCGCTCCCT	1037	                   	960	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1038	GCAGTCCCT	1039	CCGCTCCCT	1040	                   	961	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1041	GCAGTCCCT	1042	CCGCTCCCT	1043	                   	962	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1044	GCAGTCCCT	1045	CCGCTCCCT	1046	                   	963	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1047	GCAGTCCCT	1048	CCGCTCCCT	1049	                   	964	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1050	GCAGTCCCT	1051	CCGCTCCCT	1052	                   	965	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1053	GCAGTCCCT	1054	CCGCTCCCT	1055	                   	966	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1056	GCAGTCCCT	1057	CCGCTCCCT	1058	                   	967	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1059	GCAGTCCCT	1060	CCGCTCCCT	1061	                   	968	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1062	GCAGTCCCT	1063	CCGCTCCCT	1064	                   	969	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1065	GCAGTCCCT	1066	CCGCTCCCT	1067	                   	970	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1068	GCAGTCCCT	1069	CCGCTCCCT	1070	                   	971	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1071	GCAGTCCCT	1072	CCGCTCCCT	1073	                   	972	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1074	GCAGTCCCT	1075	CCGCTCCCT	1076	                   	973	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1077	GCAGTCCCT	1078	CCGCTCCCT	1079	                   	974	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1080	GCAGTCCCT	1081	CCGCTCCCT	1082	                   	975	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1083	GCAGTCCCT	1084	CCGCTCCCT	1085	                   	976	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1086	GCAGTCCCT	1087	CCGCTCCCT	1088	                   	977	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1089	GCAGTCCCT	1090	CCGCTCCCT	1091	                   	978	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1092	GCAGTCCCT	1093	CCGCTCCCT	1094	                   	979	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1095	GCAGTCCCT	1096	CCGCTCCCT	1097	                   	980	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1098	GCAGTCCCT	1099	CCGCTCCCT	1100	                   	981	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1101	GCAGTCCCT	1102	CCGCTCCCT	1103	                   	982	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1104	GCAGTCCCT	1105	CCGCTCCCT	1106	                   	983	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1107	GCAGTCCCT	1108	CCGCTCCCT	1109	                   	984	CGCTGCTGCGACCTGCGAGTCCCTAGGCTGGC	1110	GCAGT
-----	--	------	---------------------------------	-----	----------------------------------	------	---------------------------------	------	---------------------------------	-----	--	------	---------------------------------	-----	-----------	------	-----------	------	---------------------------------	-----	--	------	---------------------------------	-----	--	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	--	------	---------------------------------	-----	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-----------	------	-----------	------	---------------------------------	-----	----------------------------------	------	-------



QY 4400 G-----GAGGTGCCCTATAAACCAAGAGAGAAATACTGAAGCACAGGGGCGGAGCA 4453  
Db 1425 aGlyGlnAspGlyProGlyPro-----MetGlyProProGlyLeuProG1 1441  
QY 4454 GACCAGACAGCAGGAGGTCTCCAAAGCACAGAGTGGCAACAAACCCGAGCTGAGC 4513  
Db 1441 yLeuLysGlyAspProGly---SerLysGlyGluLysGlyHisProGlyLeuLeGlyLe 1460  
QY 4514 ATCAGGACCTTGCCTCGAATTGCTTCCAGTATTACGGTGCCTCTTCTCTGCCCCCTTTC 4573  
Db 1460 uLeGlyProProGlyGluGlnGly-----G1 1469  
QY 4574 CCAGGTATCTGTGGTTGCCAGGTGGGGAGGCGCAACCATAGCACACACAGGATTTTC 4633  
Db 1469 uLysGlyAspArgGlyLeuProGlyThrGlnGlySerPro-----1482  
QY 4634 CTGAAGTTTACATGACAGTACATTTGGGGTGTAGGT---GGCAGCTCCCAAGGCC 4690  
Db 1483 -----GlyAlaLysGlyAspGlyGlyLeProGlyPr 1493  
QY 4691 CTGC-----CCCGCAGCCACCC 4709  
Db 1493 oAlaGlyProLeuGlyProProGlyProPro 1503

## RESULT 14

US-09-735-367B-3  
; Sequence 3, Application US/09735367B  
; Patent No. US20020151477A1  
; GENERAL INFORMATION:  
; APPLICANT: Gustafsson, Jan-Ake  
; APPLICANT: Cairra, Francoise  
; APPLICANT: Antonsson, Per  
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR  
; FILE REFERENCE: 102093-100  
; CURRENT APPLICATION NUMBER: US/09/735,367B  
; CURRENT FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: US 60/174,544  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 2005  
; TYPE: PRT  
; ORGANISM: mammal  
US-09-735-367B-3

Alignment Scores:  
Pred. No.: 6,6e-06 Length: 2005  
Score: 230.50 Matches: 387  
Percent Similarity: 32.03% Conservative: 212  
Best Local Similarity: 20.70% Mismatches: 619  
Query Match: 2.44% Indels: 654  
DB: 10 Gaps: 90

US-09-931-704-3 (1-5087) x US-09-735-367B-3 (1-2005)

QY 15 CTGGCGGATGGATATTAAAGCTTCGCGGAGCGCGCTCGCCTCCCACTCCGCCA 74  
Db 257 ProProGlyTrpAsn-----GlnLeuProSerGlyAlaLeuGlnPro-----ProPro 272  
QY 75 GCTCCCGG-----AGAGGAGCGCACCCCGC-----CGGCCCGAGCCCC 113  
Db 273 AlaGlnGlySerLeuGlyThrMetThrAlaAsnGlnGlyTrpLysAlaProLeuPro 292  
QY 114 AGCCCATGAGCTCCGAGCAGGTGAAACCCAACTAGCCTGTCTTTCATACATGA 173  
Db 293 GlyProMetGlnGlnGlnLeuGlnAlaArgProSerLeuAlaThrVal-----308  
QY 174 CAAGCAGCGCCCATCTGATACCTAAACCGACCAAGTCAAGCCCTCCCACTCACCTCT 233  
Db 309 -----Gln-ThrProSerHisProProProProProPh 320

QY 234 GCTGCCAG-----243  
Db 320 eGlySerGlnGlnAlaSerGlnAlaHisThrAsnPheProGlnMetSerAsnProGlyG1 340  
QY 244 -----ACCTCACCATCTTGTGGACTCAAA-----270  
Db 340 nPheThrAlaProGlnMetLysSerLeuGlnGlyGlyProSerArgValProThrProLe 360  
QY 271 -CCTCAACCCGCTATAATCAACCAATCCCAAGTCTAAACTAATCTGAAACTTTTAAAGT 329  
Db 360 uGlnGlnProHisLeuThrAsnLysSerProAla-----SerSerProSerSerPh 377  
QY 330 AACCCAGTCTTAAACCTACCTAGCCCAATGCCAATATATATATACCTAGCCAAACCT 389  
Db 377 eGlnGlnGlySerProAlaSerSerProThrValAsn-----GlnThrG1 392  
QY 390 AACTGCCCTTTGCCAGTCCAAAGTCTCCACTGAATCCTCACCTTGGTCTCCTCACTGAAATC 449  
Db 392 nGlnGlnMet---GlyProArgProProGlnAsnAsnProLeu---ProGln---GlyPh 409  
QY 450 CCAGAAAGCATATTTCCCACTGCCCACTCCCTCTTACAGCACCCCAACCTGGCCTC 509  
Db 409 eGlnGlnProValSerSerProGlyArgAsnProMetValGlnGlnGlyAsnValProPr 429  
QY 510 TGGACTCTCGTATCTCTGGGATGTCAAACTCTGAGTGCCATCAGCCAAAGCCCGAC 569  
Db 429 oAsnPheMetVal-----MetGlnGlnProProAsnGlnGlyPr 443  
QY 570 TCGTCAATGCACCTCTCT-----CCCTTCTGTCTCCCACTTTCGAGGCTGA 617  
Db 443 oGlnSerLeuHisProGlyLeuGlyGlyMetProLysArgLeuProProGlyPheSerAl 463  
QY 618 TGAAGAGCGCTCATTTGAAGTCCAACTTTTCCCACTAACCAAGAAAGCGG---GTGAA 674  
Db 463 aGlyGlnAlaAsn-----ProAsnPheMetGlnGlyGlnValPr 476  
QY 675 CTTCCACACTGCCACCTGCTCCCTGAGAGTGAGCACTAAATCTCTCAATTAACCCAC 734  
Db 476 oSerThrThrAlaThrThrProGlyAsnSerGlyAlaProGlnLeuGlnAlaAsnGlnAs 496  
QY 735 CTTACACTTCCACACTCAGGAATCAGATCTCTAGATATATACCAAACTAAG-----786  
Db 496 nValGln-----HisAlaGlyGlyGlnGlyAlaGlyProProGlnAsnGlnMetGlnVa 514  
QY 787 -CCCATAAAGCGACCGACCTAGTGTCTAACCCCTATACCTTCTCTCTCTGAGTGTAG 845  
Db 514 lSerHis---GlyProProAsnMet-----MetGlnPro 524  
QY 846 TCTGTTCTTGGC-----GGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 899  
Db 525 SerLeuMetGlylLeHisGlyAsnMetAsnAsnGlnGlnAlaGlyThrSerGlyValPro 544  
QY 900 CAGCTGCCAGCTCTGACATGTGTCTCCACCTCTGACTCCCTCCCTCAAGCTGACGTG 959  
Db 545 GlnValAsnLeuSerAsnMetGlnGlyGlnProGlnGlnGlyProProSerGlnLeuMet 564  
QY 960 GGACTG---GAAGACTGCGAAGTAGGTACAACTGGAACACAGGAGGTTCGACCTG 1016  
Db 565 GlyMetHisGlnGlnLeValProSerGlnGlnMetValGlnGlnGlnGlnGlnGlnGln 584  
QY 1017 CAGTCCC-----TAGCCCTGGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1055  
Db 584 uAsnProGlnAsnProMetlLeLeuSerArgAlaGlnLeuMetProGlnGlnMetMe 604  
QY 1056 TATACATGTTGGGCACACACAGTGGGCACATGTCACCAAGACTCTCTCAGCTGACACACA 1115  
Db 604 tValAsnProProSerGlnAsnLeuGlyProSerProGlnArgMetThrProProLysG1 624  
QY 1116 GATCCATCTCAAGTATCTACTGATAGACACTCATGCTGCGCCAGCTCTCTCAT-----1167  
Db 624 nMetLeuSerGlnGlnGlyProGln-----MetMetAlaProHisAsnGlnMe 640  
QY 1168 -----CCTCAAAACATACATGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1220





```
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735,367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2063
; TYPE: PRT
; ORGANISM: Human
; US-09-735-367B-2

Alignment Scores:
Pred. No.: 1,18e-05 Length: 2063
Score: 226.00 Matches: 387
Percent Similarity: 32.03% Conservative: 212
Best Local Similarity: 20.70% Mismatches: 619
Query Match: 2.40% Indels: 654
DB: 1.0 Gaps: 90

US-09-931-704-3 (1-5087) x US-09-735-367B-2 (1-2063)
Qy 15 CTGCGGATGGATTATTAAGCTTCGGCGGAGCCGGCTCGCCCTCCCACTCCGCCA 74
||| |||||:|:| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 315 ProProGlyTrpAsn-----GlnLeuProSerGlyAlaLeuGlnPro-----ProPro 330
Qy 75 GCCTCCGG-----AGAGGAGCGGACCCGGC-----CGGCCAGGCC 113
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 331 AlaGlnGlySerLeuGlyThrMetThrAlaAsnGlnGlyTrpLysAlaProLeuPro 350
Qy 114 AGCCCATCGAGCTCCGAGCGAGTTGAAACCAACCAATAGCCCTGCTCTTCAATACATGA 173
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 351 GlyProMetGlnGlnLeuGlnAlaAaProSerLeuAlaThrVal----- 366
Qy 174 CAAGCAGCGCCCATCTGATACCTAAACCGACCAAGTCAAGCCCTCCCACTCACTCT 233
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 367 -----Gln-ThrProSerHisProProProProProPh 378
Qy 234 GCCTGCCAG----- 243
Db 378 eGlySerGlnGlnAlaSerGlnAlaHisThrAsnPheProGlnMetSerAsnProGlyG 398
Qy 244 -----ACCTCACCACATCTTGTGGACTCAA----- 270
Db 398 nPheThrAlaProGlnMetLysSerLeuGlnGlyGlyProSerArgValProThrProLe 418
Qy 271 -CCTCAACCGCACTAAATCAACCAATCCCAAGTCTAAACTTAATCTGAACCTTTAAAGT 329
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 uGlnGlnProHisLeuThrAsnLysSerProAla-----SerSerProSerSerPh 435
Qy 330 AACCAGTCTTAAACCTAACCTAGCCCAATGCCAATTATATCTACCTAGCCCAACCT 389
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 eGlnGlnGlySerProAlaSerSerProThrValAsn-----GlnThrG 450
Qy 390 AACTGCTTTGGCAGTCCAAAGTGTCCATCTCACTTGGTCTCCTCACTCACTGAAATC 449
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 450 nGlnGlnMet---GlyProArgProProGlnAsnAsnProLeu---ProGln---GlyPh 467
Qy 450 CCAGAAAGCATATTTCCCACTGCCACATCCCTCTTACAGCACCCACCTGGCCTC 509
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 467 eGlnGlnProValSerSerProGlyArgAsnProMetValGlnGlnGlyAsnValProPr 487
Qy 510 TGGACTCTCTGGTATCTGGGATGTCACAACTCTGCGAGTGCATCAGCCCAACAGCCGAC 569
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 oAsnPheMetVal-----MetGlnGlnGlnProProAsnGlnGlyPr 501
Qy 570 TCGTCAATATGCACCTCTCT-----CCCTTCTCTGCTCCCACTTGCAGGTGA 617
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 501 oGlnSerLeuHisProGlyLeuGlyGlyMetProLysArgLeuProProGlyPheSerAl 521
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 618 TGAAGAGGCTCATTAAGTCCAACTTTTCCCACTTCCCACTCAACCAAGAACCGG---GTGAA 674
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:50:02 ; Search time 51.6061 Seconds  
(without alignments)  
5800.643 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 9432

Sequence: 1 aacctgcagtggtgctggc.....ccttgctaaagtcttctcca 5087

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlip

-MODBL=frame+ n2p.model -DB=us-09-931-704-3  
-Q=/cgn2\_1/USPTO spo1/US0931704/runat\_27012003\_154127\_3671/app query.fasta\_1.7189  
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0931704 @CGN 1.1.89 @runat\_27012003\_154127\_3671 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/iaa/5A COMB.pcp.\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB.pcp.\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pcp.\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pcp.\*  
5: /cgn2\_6/ptodata/1/iaa/6CTUS COMB.pcp.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	896	9.5	225	1	US-08-792-019B-2
2	896	9.5	225	1	US-09-106-182-2
3	896	9.5	225	3	US-08-988-819-2
4	896	9.5	225	4	US-09-016-534-2
5	863	9.1	225	1	US-08-792-019B-5
6	863	9.1	225	3	US-08-988-819-5
7	863	9.1	225	4	US-09-016-534-5
C 8	330.5	3.5	1078	3	US-08-963-825-21
C 9	330.5	3.5	1078	4	US-09-500-811-21
C 10	330.5	3.5	1078	4	US-09-570-573-21
C 11	330.5	3.5	1078	4	US-09-548-608-21
C 12	327.5	3.5	1057	3	US-08-931-820-4

13	318.5	3.4	2972	4	US-09-579-181-2	Sequence 2, Appli
14	318.5	3.4	3118	4	US-09-579-181-1	Sequence 1, Appli
C 15	305	3.3	3418	3	US-08-963-825-20	Sequence 20, Appl
C 16	305	3.3	3418	4	US-09-010-999-1	Sequence 1, Appli
C 17	305	3.3	3418	4	US-09-500-811-20	Sequence 20, Appl
C 18	305	3.3	3418	4	US-09-570-573-20	Sequence 20, Appl
C 19	305	3.3	3418	4	US-09-548-608-20	Sequence 20, Appl
C 20	302	3.2	1060	3	US-08-931-820-3	Sequence 3, Appli
C 21	298	3.2	870	2	US-09-010-928B-2	Sequence 2, Appli
C 22	296	3.1	1064	1	US-08-642-255-62	Sequence 62, Appl
C 23	295.5	3.1	1185	4	US-09-041-886-23	Sequence 23, Appl
C 24	295.5	3.2	1366	3	US-08-963-825-19	Sequence 19, Appl
C 25	295.5	3.2	1366	4	US-09-500-811-19	Sequence 19, Appl
C 26	295.5	3.2	1366	4	US-09-570-573-19	Sequence 19, Appl
C 27	295.5	3.2	1366	4	US-09-548-608-19	Sequence 19, Appl
C 28	295	3.2	745	2	US-09-010-928B-28	Sequence 28, Appl
C 29	295	3.2	1442	2	US-08-316-650-12	Sequence 12, Appl
C 30	295	3.2	1442	5	PCT-US95-02251-12	Sequence 12, Appl
C 31	293.5	3.1	1024	3	US-08-931-820-2	Sequence 2, Appli
C 32	293.5	3.1	1461	4	US-09-585-887-9	Sequence 9, Appli
C 33	293.5	3.1	1461	4	US-09-289-578-9	Sequence 9, Appli
C 34	291	3.1	1057	3	US-08-931-820-1	Sequence 1, Appli
C 35	289	3.1	822	4	US-09-219-849-49	Sequence 49, Appl
C 36	288	3.1	1958	1	US-07-945-283-2	Sequence 2, Appli
C 37	287	3.0	960	4	US-09-219-849-5	Sequence 5, Appli
C 38	286	3.1	1366	4	US-09-585-887-10	Sequence 10, Appl
C 39	286	3.1	1366	4	US-09-289-578-10	Sequence 10, Appl
C 40	279.5	3.0	897	1	US-08-397-633A-50	Sequence 50, Appl
C 41	278	3.0	1341	3	US-08-963-825-18	Sequence 18, Appl
C 42	278	3.0	1341	4	US-09-500-811-18	Sequence 18, Appl
C 43	278	3.0	1341	4	US-09-570-573-18	Sequence 18, Appl
C 44	278	3.0	1341	4	US-09-548-608-18	Sequence 18, Appl
C 45	277.5	3.0	1065	1	US-08-642-255-72	Sequence 72, Appl

#### ALIGNMENTS

RESULT 1  
US-08-792-019B-2  
; Sequence 2, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUROTROPIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/792.019B  
; APPLICATION NUMBER: US/08/792.019B  
; FILING DATE: 03-FEB-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-792-019B-2

Alignment Scores:  
Pred. No.: 6,41e-57 Length: 225  
Score: 896.00 Matches: 170  
Percent Similarity: 90.72% Conservative: 6  
Best Local Similarity: 87.63% Mismatches: 6  
Query Match: 9.50% Indels: 12  
DB: 1 Gaps: 2

US-09-931-704-3 (1-5087) x US-08-792-019B-2 (1-225)

QY 3311 GGTGAGCAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349  
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51  
QY 3350 -----CTCTCCTTTTCACAGCTGAACTACTGGGCCCCCTTTCAACGAG 3394  
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71  
QY 3395 CCAGACTTCAACCTCCCGCTGGGGGCGAGAGACTCTGCCAGGGCCACTGTGTGACTTG 3454  
Db 72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91  
QY 3455 GAGGTGTGGCGAAGCTCAATACAACTGGCGTGAACCCAGAGACTACGAGGCTACAGC 3514  
Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111  
QY 3515 CACCTTCTGTGTACTTGTGGTGGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGCCGC 3574  
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131  
QY 3575 AGCCTGGCCCACTTCTGCACCGCTCCAGGGCTCTGGGCGAGACTTCTGGCTGCTGAAGGAGCTG 3634  
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151  
QY 3635 GCAGCTCTGGGTACCCACTGCCAGCGCTGCCCTGGGAGTCAACCCACTTGGACTCCT 3694  
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171  
QY 3695 GCGCTGCGCCACAGTACTTCTCCAGAGAGATGGACGACTTCTGGCTGCTGAAGGAGCTG 3754  
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191  
QY 3755 CAGACTGCTGGCGCTCGGCCAAGAGACTTCAACCGCTCAAGAGAGAGTGCAGCCT 3814  
Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211  
QY 3815 CCAGAGCTGCAGTACCTGACCTGGGGCTCATGGCTTC 3856  
Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 2  
US-09-106-182-2  
; Sequence 2, Application US/09106182  
; Patent No. 6046035  
; GENERAL INFORMATION:  
; APPLICANT: Shi, Yangu  
; APPLICANT: Ruben, Steve  
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc  
; STREET: 9410 Key West Ave  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: US  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09106,182

FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/051,053  
FILING DATE: 30-JUN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF385  
TELEPHONE: 301-309-8504  
TELEFAX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-106-182-2

Alignment Scores:  
Pred. No.: 6,41e-57 Length: 225  
Score: 896.00 Matches: 170  
Percent Similarity: 90.72% Conservative: 6  
Best Local Similarity: 87.63% Mismatches: 6  
Query Match: 9.50% Indels: 12  
DB: 1 Gaps: 2

US-09-931-704-3 (1-5087) x US-09-106-182-2 (1-225)

QY 3311 GGTGAGCAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349  
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51  
QY 3350 -----CTCTCCTTTTCACAGCTGAACTACTGGGCCCCCTTTCAACGAG 3394  
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71  
QY 3395 CCAGACTTCAACCTCCCGCTGGGGGCGAGAGACTCTGCCAGGGCCACTGTGTGACTTG 3454  
Db 72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91  
QY 3455 GAGGTGTGGCGAAGCTCAATACAACTGGCGTGAACCCAGAGACTACGAGGCTACAGC 3514  
Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111  
QY 3515 CACCTTCTGTGTACTTGTGGTGGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGCCGC 3574  
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131  
QY 3575 AGCCTGGCCCACTTCTGCACCGCTCCAGGGCTCTGGGCGAGACTTCTGGCTGCTGAAGGAGCTG 3634  
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151  
QY 3635 GCAGCTCTGGGTACCCACTGCCAGCGCTGCCCTGGGAGTCAACCCACTTGGACTCCT 3694  
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171  
QY 3695 GCGCTGCGCCACAGTACTTCTCCAGAGAGATGGACGACTTCTGGCTGCTGAAGGAGCTG 3754  
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191  
QY 3755 CAGACTGCTGGCGCTCGGCCAAGAGACTTCAACCGCTCAAGAGAGAGTGCAGCCT 3814  
Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211  
QY 3815 CCAGAGCTGCAGTACCTGACCTGGGGCTCATGGCTTC 3856  
Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

```

; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-988-819-2

Alignment Scores:
Pred. No.: 6.41e-57 Length: 225
Score: 896.00 Matches: 170
Percent Similarity: 90.72% Conservative: 6
Best Local Similarity: 87.63% Mismatches: 6
Query Match: 9.50% Indels: 12
DB: 3 Gaps: 2

US-09-931-704-3 (1-5087) x US-08-988-819-2 (1-225)
QY 3311 GGTGAGCCAGCGCTGCGCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCCTTTTCACAGCTGAACCTACCTGGCGCCCTTTCACAGCAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCAACCCCTCCCGCTGGGGCAGAGACTTGCCAGGGCCACTGTGACTTG 3454
Db 72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91
QY 3455 GAGGTGTGGCGAAGCTCATGATCAAACTCGCGCTGACCCAGAACTACGAGGCGCTACG 3514
Db 92 GluValTrrArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTGTTGCTGGCTCAACCGTCAGGCTCCACTGCTGAGCTGCGCGC 3574
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgArg 131
QY 3575 AGCTGGCCCACTTCTGCACAGCCCTCCAGGGCTGCTGGGCGAGCATTTGGGGCGTCATG 3634
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGTACCCACTGCGCCCGAGCGCTGCTGGGAGCTGAACCCACTTGAGCTCT 3694
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171

QY 3695 GGCCTCCACACAGTACTCTCTCCAGAGATGGACGACTTCTTGCTGTGAAGAGCTG 3754
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191
QY 3755 CAGACCTGGCTGTGGCGCTCGCCCAAGGACTTCAACCGGCTCAAGAGAGATGACAGCT 3814
Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211
QY 3815 CCAGCAGTGCAGTCACCTGCACCTGCACCTGGGGCTCATGGCTC 3856
Db 212 ProAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 4
US-09-016-534-2
; Sequence 2, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULIA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,534
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-016-534-2

Alignment Scores:
Pred. No.: 6.41e-57 Length: 225
Score: 896.00 Matches: 170
Percent Similarity: 90.72% Conservative: 6
Best Local Similarity: 87.63% Mismatches: 6
Query Match: 9.50% Indels: 12
DB: 4 Gaps: 2

US-09-931-704-3 (1-5087) x US-09-016-534-2 (1-225)
QY 3311 GGTGAGCCAGCGCTGCTCCCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCCTTTTCACAGCTGAACCTACCTGGCGCCCTTTCACAGCAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
```

QY 3395 CCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGGCCACTGTGACTTG 3454  
Db |||||  
72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91  
QY 3455 GAGGTGTGGGAAGCTCAATGACAACTCGCGCTGACCCAGAACTACGAGGCTTACAGC 3514  
Db |||||  
92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111  
QY 3515 CACCTTCTGTGTACTTGGTGGCTCAACCGTCAGGCTCCACTGCTGAGCTGCGCGC 3574  
Db |||||  
112 HisGlnLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131  
QY 3575 AGCTGGCCCACTTCTGCACCACTCCAGGCTCTGCTGGGCGACATTCGGGGCTCATG 3634  
Db |||||  
132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuGlySerIleAlaGlyValMet 151  
QY 3635 GCAGCTCTGGGTACCACTGCGCCCGCTGCTGGGAGTGAACCCACTTGGACTCCT 3694  
Db |||||  
152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpPro 171  
QY 3695 GGCCTTGGCCACAGTCTCTCCAGAGATGAGCACTTCTGGCTGCTGAAGGAGCTG 3754  
Db |||||  
172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191  
QY 3755 CAGACTGTGCTGGCGCTGGCCAAAGGACTTCAACCGGCTCAAGAAAGATGCAAGCT 3814  
Db |||||  
192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211  
QY 3815 CCAGAGCTGCAGTCACTCCCTGACCTGGGGCTCATGGCTTC 3856  
Db |||||  
212 ProAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

## RESULT 5

US-08-792-019B-5  
; Sequence 5, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/792,019B  
; FILING DATE: 03-FEB-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-792-019B-5  
Alignment Scores:  
Pred. No.: 1,55e-54 Length: 225  
Score: 863.00 Matches: 163  
Percent Similarity: 86.66% Conservative: 9  
Best Local Similarity: 84.02% Mismatches: 10

Query Match: 9.15% Indels: 12  
DB: 1 Gaps: 2  
US-09-931-704-3 (1-5087) x US-08-792-019B-5 (1-225)  
QY 3311 GGTGAGCCAGCGGTCTGCTCCCTCC-----TCCTTCCCATCACC----- 3349  
Db |||||  
32 GlyAspProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51  
QY 3350 -----CTCTCCTTTTCACAGCTGAACTACCTGGGCCCCCTTTCAACGAG 3394  
Db |||||  
52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71  
QY 3395 CCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGGCCACTGTGACTTG 3454  
Db |||||  
72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAsnLeu 91  
QY 3455 GAGGTGTGGGAAGCTCAATGACAACTCGCGCTGACCCAGAACTACGAGGCTTACAGC 3514  
Db |||||  
92 GluValTrpArgSerLeuAsnAspArgLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111  
QY 3515 CACCTTCTGTGTACTTGGTGGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGCGC 3574  
Db |||||  
112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131  
QY 3575 AGCTTGGCCCACTTCTGCACCACTCCAGGCTCTGCTGGGCGACATTCGGGGCTCATG 3634  
Db |||||  
132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuGlySerIleAlaGlyValMet 151  
QY 3635 GCAGCTCTGGGTACCACTGCGCCCGCTGCTGGGAGTGAACCCACTTGGACTCCT 3694  
Db |||||  
152 AlaThrLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProAlaTrpAlaPro 171  
QY 3695 GGCCTTGGCCACAGTCTCTCCAGAGATGAGCACTTCTGGCTGCTGAAGGAGCTG 3754  
Db |||||  
172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191  
QY 3755 CAGACTGTGCTGGCGCTGGCCAAAGGACTTCAACCGGCTCAAGAAAGATGCAAGCT 3814  
Db |||||  
192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211  
QY 3815 CCAGAGCTGCAGTCACTCCCTGACCTGGGGCTCATGGCTTC 3856  
Db |||||  
212 ProAlaAlaSerValThrLeuHisLeuGluAlaHisGlyPhe 225

## RESULT 6

US-08-988-819-5  
; Sequence 5, Application US/08988819  
; Patent No. 6054294  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/988,819  
; FILING DATE: 12-DEC-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:

; NAME: COOK, ROBERT R.  
 ; REGISTRATION NUMBER: 31,602  
 ; REFERENCE/DOCKET NUMBER: A-442A  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 225 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-988-819-5

Alignment Scores:  
 Pred. No.: 1,55e-54 Length: 225  
 Score: 863.00 Matches: 163  
 Percent Similarity: 88.66% Conservative: 9  
 Best Local Similarity: 84.02% Mismatches: 10  
 Query Match: 9.15% Indels: 12  
 DB: 3 Gaps: 2

US-09-931-704-3 (1-5087) x US-08-988-819-5 (1-225)

```

QY 3311 GGTGAGCCAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCTTTTTCACAGCTGAACACTACCTGGGCCCCCTTTCAACGAG 3394
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCACACCTCCCGGCTGGGGGAGAGACTCTGCCAGGGCCACTGTGTGACTTG 3454
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAsnLeu 91
QY 3455 GAGGTGTGGGAAGCCTCAATGACAACTCGGCTGACCCAGAGACTACGAGGCTACAGC 3514
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92 GluValTrpArgSerLeuAsnAspArgLeuGluThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGTGGCTCAACCGCTGCGGACTGACCTGTGAGCTGCGCGC 3574
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeuArgArg 131
QY 3575 AGCGTGGCCACTTCTGTCACAGCCCTCCAGGCGCTGCTGGGCGAGCATTCGGGCGTCATG 3634
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGCTACCACTGCCAGCGCTGCTGGGACTGAACCCACTTGGACTCCT 3694
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 AlaThrLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProAlaTrpAlaPro 171
QY 3695 GGCCTGCCCCACAGTACTTCTCCAGAGATGAGACTTCTGGCTGTGAAGAGAGCTG 3754
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLysGluLeu 191
QY 3755 CAGACTGTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 3814
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211
QY 3815 CCAGAGCTGCACTACCCCTGACCTGGGCGCTGCTGGGCTGCTGGGCTGCTGGGCTGCT 3856
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 ProAlaAlaSerValThrLeuHisLeuGluAlaHisGlyPhe 225
  
```

# RESULT 7

US-09-016-534-5  
 ; Sequence 5, Application US/09016534  
 ; Patent No. 6143874  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHANG, MING-SHI  
 ; APPLICANT: ELLIOTT, GARY S.  
 ; APPLICANT: SARMIENTO, ULLA  
 ; APPLICANT: SENALDI, GIORGIO  
 ; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
 ; NUMBER OF SEQUENCES: 16  
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: AMGEN INC.  
 ; STREET: ONE AMGEN CENTER  
 ; CITY: THOUSAND OAKS  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 91320  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/016,534  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/792,019  
 ; FILING DATE: 03-FEB-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: COOK, ROBERT R.  
 ; REGISTRATION NUMBER: 31,602  
 ; REFERENCE/DOCKET NUMBER: A-442B  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 225 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-09-016-534-5

Alignment Scores:  
 Pred. No.: 1,55e-54 Length: 225  
 Score: 863.00 Matches: 163  
 Percent Similarity: 88.66% Conservative: 9  
 Best Local Similarity: 84.02% Mismatches: 10  
 Query Match: 9.15% Indels: 12  
 DB: 4 Gaps: 2

US-09-931-704-3 (1-5087) x US-09-016-534-5 (1-225)

```

QY 3311 GGTGAGCCAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCTTTTTCACAGCTGAACACTACCTGGGCCCCCTTTCAACGAG 3394
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCACACCTCCCGGCTGGGGGAGAGACTCTGCCAGGGCCACTGTGTGACTTG 3454
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAsnLeu 91
QY 3455 GAGGTGTGGGAAGCCTCAATGACAACTGCGGCTGACCCAGAGACTACGAGGCTACAGC 3514
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92 GluValTrpArgSerLeuAsnAspArgLeuGluThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGTGGCTCAACCGCTGCGGACTGACCTGTGAGCTGCGCGC 3574
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeuArgArg 131
QY 3575 AGCGTGGCCACTTCTGTCACAGCCCTCCAGGCGCTGCTGGGCGAGCATTCGGGCGTCATG 3634
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGCTACCACTGCCAGCGCTGCTGGGACTGAACCCACTTGGACTCCT 3694
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 AlaThrLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProAlaTrpAlaPro 171
QY 3695 GGCCTGCCCCACAGTACTTCTCCAGAGATGAGACTTCTGGCTGTGAAGAGAGCTG 3754
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLysGluLeu 191
QY 3755 CAGACTGTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 3814
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211
QY 3815 CCAGAGCTGCACTACCCCTGACCTGGGCGCTGCTGGGCTGCTGGGCTGCTGGGCTGCT 3856
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 ProAlaAlaSerValThrLeuHisLeuGluAlaHisGlyPhe 225
  
```

Db 192 GlnThrTyrLeuTipArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211  
Qy 3815 CCAGAGCTGCAGTACCTGCCTGCGGGGCTCATGGCTTC 3856  
Db 212 ProAlaAlaSerValThrLeuHisLeuGluAlaHisGlyPhe 225

## RESULT 8

US-08-963-825-21  
; Sequence 21, Application US/08963825  
; Patent No. 6110689  
; GENERAL INFORMATION:  
; APPLICANT: Ovis, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,825  
; FILING DATE:  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/187,319  
; FILING DATE: 21-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1078 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (III)  
US-08-963-825-21

Alignment Scores:  
Pred. No.: 7.69e-16 Length: 1078  
Score: 330.50 Matches: 319  
Percent Similarity: 29.53% Conservative: 82  
Best Local Similarity: 23.49% Mismatches: 502  
Query Match: 3.54% Indels: 456  
Gaps: 71

US-09-931-704-3 (1-5087) x US-08-963-825-21 (1-1078)

Qy 5059, GGGAACTGGGTGGACAGAGTGTGTGGAGGACTGAGGGATCAAGGGTTGGTGGT 5000  
Db 69 GlyProProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly 84  
Qy 4999 GGCAGCAGCAGTGGTAAGTCTGAGTGTGCGCCCATCGCCCGGAGGAGAGGG 4940

Db 85 GluSer-----GlyArgProGlyArgPro-----GlyAspArgGly 96  
Qy 4939 ---CCTCTTCCAGAGCTCTGTGTGCTGCCACACAGAGCAAGAGGGAGTGG----- 4892  
Db 97 LeuProGlyProProGlyIle-----LysGlyProAlaGlyIleProGlyPhe 112  
Qy 4891 ---GGAGCCAGGGGTGGCAAGAGTCTGTATGATGACACAGAGGGGTGAGCTGGGCAAGA 4835  
Db 113 ProGlyMetLysGlyHisArgGly---PheAspGlyArgAsnGlyGluLysGlyGlu--- 130  
Qy 4834 GACTAAGTCTGGGGCAAGCAATTTTATTGTATACAGAATAGAAATCTGCATATAA 4775  
Db 130 ----- 130  
Qy 4774 TATCATCTAATAATAACATCTCCAAATAATAATAATAATAATAACACACACTTAGATCA 4715  
Db 130 ----- 130  
Qy 4714 TGAGTGGGTGGGGCTGGGGGAGGGCCCTGGGGAGCTGCCA----- 4673  
Db 131 ---ThrGlyAlaProGlyLeuLysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaPro 149  
Qy 4672 ---CCCTACACCCCAAAATGCTACTGCAATTGTAACCTTCAGGAAATCCTGTGGTGTGCG 4616  
Db 150 GlyProMetGlyProArg----- 155  
Qy 4615 TATGGTTGCCCTCCAGCCTGGCAACCCACAGATACCTCGGAAAGGGGAGAGAGA 4556  
Db 156 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg--- 173  
Qy 4555 GGCACCGTAATCTCGAAGCAATTCGAGGCAAGTCTGATGCTCAGCTCGGTTTGT 4496  
Db 174 -----GlyAsnAspGlyAlaArgGlySerAspGlyGlnProGlyProPro 188  
Qy 4495 TTGCCACTCTGTCTTTGGAGACTCTCGGTCTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 4436  
Db 189 GlyProProGlyThrAlaGlyPheProGly----- 198  
Qy 4435 TTCAGTATTTCTCTCTGTGTGTATAGGGCACCTCCAATGCAGACCCCAACCTGCACGTG 4376  
Db 199 -----SerProGlyAlaLysGlyGluValGlyProAlaGlySerPro----- 212  
Qy 4375 TTAATCTCATCTCTTTTGTGTAGAATTGCCACCAATGTTTGTCTGTGAATGGATGTAT 4316  
Db 212 ----- 212  
Qy 4315 AGAGTATTGGGATAGGAGAGCAACCCATCCAAAGAGAGTGGGGGAGAGAGGC 4256  
Db 213 -----GlySerAsnGlyAlaProGlyGlnArgGlyGluPro----- 224  
Qy 4255 TTCACCCAGAAAGGCATGTGTGTGGTGGCCCTCAACCCCTGCCTCCAGATGTGCCA 4196  
Db 225 ---GlyProGlnGlyHisAlaGly-----AlaGlnGlyProProGly----- 237  
Qy 4195 CCTGAGGGGCTGGGTGGGCAAGTGTAGTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4136  
Db 238 ProProGlyIleAsnGlySer---ProGlyGlyLysGlyGluMetGly----- 252  
Qy 4135 CTCTGCACCAACCTGAACCACTTCACATCTCCTCAGCATGACTTCCTGTAGAAACAGGA 4076  
Db 253 -----ProAlaGlyIleProGlyAla-----ProGlyLeuMetGly 264  
Qy 4075 CAGGACAGGGCTACTGTACTCTCTCCCTCCAGCTCGGTAGACCT-----TTGGGAGGTGG 4022  
Db 265 -AlaArgGlyProProGlyProAlaGlyAlaAsnGlyAlaProGlyLeuArgGlyGlyAl 284  
Qy 4021 GGAGGAGAGAGGGTGTATCGCATCACGCCCGCCGCTCCAGGAAAGGGCCAGAGCTC 3962  
Db 284 aGlyGluProGlyLysAsnGlyAlaLysGlyGluProGlyProArgGlyGluArgGlyGly 304  
Qy 3961 ACAGTTCTGTCT-----CCTGGCTCAACAGGTGT---GGCATACAGGGCTGGCTCTC 3911



Db 304 uAlaGlyIleProGlyValProGlyAlaLysGlyGluAspGlyLysAspGly-SerProG 324  
QY 3910 ACAAGTGGAGCAGGTTTGAAGGGGAGCGCAAGAGAGAGAGTTCAGAAAGTCAGAAAGCC 3851  
Db 324 LyAspProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyAlaLeuGlySerA 344  
QY 3850 ATGAGCCCCAGAGTGCAGGTGACTGCAGCTGCTGGAGGCTGCATCTTCTTCTTGGAGCG 3791  
Db 344 rGlyProAlaGly- 348  
QY 3790 GTTGAAGTCTTGGCCGAGCCACAGCCAGGTCTGCAGCTCTGCAGCTCTCAGCAGCCAGATC 3731  
Db 349 -----ProAsnGlyIleProGlyGluLysGlyProAlaGlyGluArgGlyA 364  
QY 3730 GTCCATCTCTGGAGGAAGTCACTGTGGGCGAGGCGCAGGAGTCCAAAGTGGGTTTCAGTCCC 3671  
Db 364 laProGlyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArg- 379  
QY 3670 AGGCAGCGGCTGGGCGAGTGGTAGCCAGCAGCTGCCATGAGCGCCGCAATGCTGCCAG 3611  
Db 380 -----AspGlyValProGlyGlyProGlyMetArgGlyMetProGly- 396  
QY 3610 CAGGCCCTGAGGCTGTGCAGAAAGTGGCGCAGGCTGGCGCAGCTCAGCAGTGGCAGC 3551  
Db 396 lyGlyProGlySerAspGlyLysProGlyPro- 406  
QY 3550 CTGACGGTTGAGGCCACGCAAGTAACACAGAAGTGGCTGTAGGCTCTGTAGTCTGGGT 3491  
Db 407 -----GlyArgProGlyProGlyProGlyProSerGlyProArgGlyG 427  
QY 3490 CAGCCGCGATTGTCTATTGAGGCTTCGCCACACACCTCCAAAGTCAACAGTGGCCCTGGGCG 3431  
Db 412 luSer- 411  
QY 3430 AGTCTCTGCCCCAGCGGGGAGGTTTGAAGTCTGGCTCGTTGAAGGGGGCCAGCTA 3371  
Db 427 ln- 411  
QY 3370 GTTCAGCTGTAAAGAGAGAGGTTGATGGGAAAGAGAGAGGAGGCGGCTGGCTCACC 3311  
Db 442 laProGlyLys-AsnGlyGluArgGlyGlyProGlyGly- 454  
QY 3310 ACCAAGGAGATACCTGCTCCCAAGTTCTATTTTGTGCTCCCTGCACATGCGCCCTGT 3251  
Db 455 -----ProGlyProGlyPro- 462  
QY 3250 CTCCATGCTAGGCTCTTGTTCATGCGCTCCCTGAGTGTTCCTCAATAGGATATCATTT 3191  
Db 463 LysAsnGlyGluTyArgGlyProGlyPro- 473  
QY 3190 GTATGAGGTGACAGTTGAAGCAGTGGCGGTGTGAGCTCTCCAATGAGAGCGGTAGAA 3131  
Db 474 -----GlyPro 475  
QY 3130 AAAGGAGGAGGAGGAGCAGCACAAGAC- -GCTCAGAAAGACGTGCTGATGCTGGGT 3074  
Db 476 ThrGlyProGlyAspLysGlyAspThrGlyProArgGlyProGlyGlyLeuGlyGly 495  
QY 3073 GGGCAAACTTGGAGAGTCCCGCTGTGGGGGAGGAGAGAGAGAGAGGAGCTTGA 3014  
Db 496 LeuProGlyThrGlyGlyProPro- -GlyGluAsnGlyLysProGlyGluProGly 513  
QY 3013 GGAGCTGGAGAGTAGAAGTGGCCAGGAGAGTGCAGAGACCAAGAAAGAGGATTCGA 2954  
Db 514 ProLysGlyGluAla- -GlyAlaProGlyAlaProGlyGlyLysGlyAspAlaGlyAla 532  
QY 2953 GAAGGAGCAGCGCTCAGCAGTGGCTGCAAAACACACAGCAGCAGAGAGATGACCTTCA 2894  
Db 533 ProGly- 539  
QY 2893 AAACACTGTGGGCT- -TGGCAAGGAGGTGCTTGCAGCCTTGGAGAGACCATTTG 2840  
Db 540 GlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyProPro- 554

QY 2839 CTGGGCGCTTGAAGCCAGATTTCAGGGGTTTAAAGAGTGGTGGTGGAGGAGTGA 2780  
Db 555 ---GlyPro- 569  
QY 2779 GCAGAGGCTGTAGACCACTCTCTTGGAGAAATTTGGCAGGGAAGCCAGGAGAGAAA- 2723  
Db 570 GlyAlaAlaGlyThrProGlyLeu- -GlnGlyMetProGlyGluArgGly 585  
QY 2722 -----RAGGTCGGGAAGTACTAGAGGAGGCGAGTGGAGTGAAGCAA 2684  
Db 586 GlyLeuGlySerProGlyProLysGlyAspLysGlyGluProGlyGlyProGlyAlaAsp 605  
QY 2683 GGTCTTCTGCACAAAGAAACAGAGAGATGAGAGTGGCCCATTTGGAACACAGAGCCCAT 2624  
Db 606 GlyValProGlyLysAspGlyProArgGlyProThrGlyProLeuGlyProGly- 624  
QY 2623 TTGCGCTGACACACCCAGGCCACCCAGCTGAGGGCTGGGAGAGACGGGTACTCTACTA 2564  
Db 625 ---ProAlaGlyGlnProGlyAspLysGlyGluGly- -GlyAlaProGlyLeuProGly 642  
QY 2563 ACACCTCCGAAAGAGGGTTCA- -GCTGATAGGAGCAGGGTT- 2525  
Db 643 IleAlaGlyProArgGlySerProGlyGluArgGlyGluThrGlyProGlyProAla 662  
QY 2524 -----GCAGAAAGAAATGAGGAGGAGTCTCATGAGTAGCAGAGGAGG 2477  
Db 663 GlyPheProGlyAlaProGlyGlnAsnGly- -GluProGlyGly- 676  
QY 2476 CTATGGAAGACAGCATGTGTATAGGGAAGGCGCAGAGCCCTTTCCAAACCTCGGGGAGC 2417  
Db 677 -----LysGlyGlu 679  
QY 2416 GSCTGGGCGCCAGGAAAGGGATGAGGAGATGTATACTGGGAAGAAAGAGAGCA 2357  
Db 680 ArgGlyAlaProGlyGlu- -LysGlyGlu 688  
QY 2356 GGGGAGAGCGCTGGGATAGGCAAGGCCCAAGGGAGGTTTCACACGCTGACAGCAGACA 2297  
Db 689 GlyGlyProGlyValAlaValProGlyGlySerGlyProAlaGly-ProProGly 708  
QY 2296 GACAACTAATTTCTCGGTGTATCTCGGGCTGGAGCAGCTCCCGCCGACCTTCTTC 2237  
Db 708 y- -ProGlnGlyValLysGlyGluArgGlySerProGlyGlyProGlyAl 724  
QY 2236 CTTTTTCTTCTCCCTCCAGCAGTAGGCTCTGGGATGTGGAGAGGAACACCGGAGAG 2177  
Db 724 alaGlyPhePro- -GlyAlaAr 731  
QY 2176 GGGG- -CCGGTTCCCATGGTGGCTCGCTCTCCACCCCTCTCTCTCTCTCTCTCTCT 2120  
Db 731 gGlyLeuProGlyProGlySerAsnGlyAsnProGly- -ProProGlyProSerGly 750  
QY 2119 CTCTCTGAACCTCTGGACCAACCCCAATCTTACAGATTAACCTCCATCCACACCT 2060  
Db 750 ySerProGlyLysAspGly- -ProProGlyProAlaGlyAsnThrGly- 765  
QY 2059 AAACCCCTGCGCACTACTGCTGGTTCCTACAGTCCCTTCCCATGACCCCTCTAG 2000  
Db 766 -----AlaProGlySerProGlyVal-SerGlyProLysGlyAspAlaG 780  
QY 1999 GAGGCCAGCTAGGATGGAGAGGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1940  
Db 780 lyGlnProGlyGluLysGlySerProGlyAlaGlnGlyProGly- -AlaProGly 798  
QY 1939 GGGCC- -CAGGTCTGGGCGCAGGAGGAGCTGGAAACGAGGAGGAGGAGGAGGAGG 1892  
Db 798 lyProLeuGlyIleAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetP 818  
QY 1891 TGGGGCCCTC- -TGTCACCAACCAACCCCGCGGCAAGAGACACACACACAGGTCG 1835  
Db 818 roGlyProArgGlySerProGlyProGlnGlyValLysGlyGluSerGlyLysProGlyA 838

```
QY 1834 TAAAC-----AGAGGAGACTTCCCGCGCCCGCCAGGGGAGCGGCGC----- 1794
Db 838 laAsnGlyLeuSerGlyGluA-TgGlyPro---ProGlyProGlnGlyLeuProGlyLeuA 857
QY 1793 --GGTGGGCTGGACCCAGTGGGGCAATGGGCAAGGGGGTGGCAGGCTCCCCACAT 1736
Db 857 laGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeuPro--- 875
QY 1735 GCCTAGTCTGGGGCCAGGATGGGAAGTGGGGGGCCACCTAAGAGGGAAGGCGAGGA 1676
Db 876 -----GlyArg---AspGlySerProGlyGlyGlyAspA 887
QY 1675 GATAGACACAGAGGGGAGTGGTGGGAGCCAAAGAGCCCTCTCCACCCATCATTTAC 1616
Db 887 Tg-----GlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisProG 903
QY 1615 CAGAACCTCCACGGTTAGGACTGTCTTCCCAACTCTCACTCTCTCCCAACTCCAG 1556
Db 903 LyProProGly-----ProValGlyProA 911
QY 1555 ATTCTACCTGACTACATAGTCCAGCCAGCTCGGAGTGGTGGTCCAGGTA 1496
Db 911 laGlyLysSerGlyAspArgGlyGluSerGlyProAlaGlyProAlaGlyAlaProGlyP 931
QY 1495 GCGGCTGAGGTATAGGTTTCTGGATGGAGGGG---CCAGGCCCTGGGTCCCTGTGC 1440
Db 931 ro-----AlaGlySerArgGlyAlaProGlyProGlnGlyProArgG 945
QY 1439 GATTGAGACTGCATCGAGGAGGTGCCAGAGCACCCTGCAGCAGCGCTTAACATCC 1380
Db 945 lyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlyIleLysGlyHisArgGlyPheP 965
QY 1379 CC-----CACGAGCTCTGTGGCGAGGATGGACGAGAAGCATG 1341
Db 965 roGlyAsnProGlyAlaProGlySerProGlyProAlaGlyGlnGlnGlyAlaIleGlyS 985
QY 1340 AGCGCGCGCCG---GGCCACACGCGGAGAGCAGCTGCGAGCCCTCAGGTGTCTGTCCCC 1284
Db 985 erProGlyProAlaGlyProArgGlyProVal-----GlyProSerGlyProGlyL 1003
QY 1283 ACCTTGGCTGGGTGGGGGTGAGGGT-----GGGACACAGATGGGAGGCGAGGGGA 1230
Db 1003 yeAspGlyThrSer-GlyHisProGlyProIleGlyProGlyProArgGlyAsnArg 1022
QY 1229 TGGAGAGGGGAAC-----ACTCTGGCAAGACGGAGAG 1194
Db 1023 GlyGluArgGlySerGluGlySerProGlyHisProGlyGln 1036

RESULT 9
US-09-500-811-21
; Sequence 21, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE: INFORMATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-09-500-811-21

Alignment Scores:
Pred. No.: 7,69e-16 Length: 1078
Score: 330.50 Matches: 319
Percent Similarity: 29.53% Conservative: 82
Best Local Similarity: 23.49% Mismatches: 502
Query Match: 3.54% Indels: 456
DB: 4 Gaps: 71

US-09-931-704-3 (1-5087) x US-09-500-811-21 (1-1078)
QY 5059 GGGAAACCTGGGGTGGAGAGAGCTGTGGGAGACTGAGGATCAAGGGGTGTGTGCT 5000
Db 69 GlyProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly 84
QY 4999 GCGAGCAGCAGTGGGTAACTCTGAGGTGCGCCATCGCCCGAGTGGCGGAGGAGAGAGG 4940
Db 85 GluSer-----GlyArgProGlyArgPro-----GlyAspArgGly 96
QY 4939 ---CCTCTTCCCAGAGCTCTGTGCAGCTCCACAGAGCAAGCAGGAGTGTG----- 4892
Db 97 LeuProGlyProProGlyIle-----LysGlyProAlaGlyIleProGlyPhe 112
QY 4891 ---GGAGCCAGGGTGGCAAGAGTCTGATGACACAGGAGGGGTGAGCTGGGCAAGA 4835
Db 113 ProGlyMetLysGlyHisArgGly---PheAspGlyArgAsnGlyGluLysGlyGlu--- 130
QY 4834 GACTAAGTTCTGGGCAAGCATTTTATTGTTTAAATACAAGATAGAAATTTCTGCAATAAA 4775
Db 130 ----- 130
QY 4774 TATCATCTAATAATAATACATCTCCAAATAATAATAATTAATACACACACTTAGATCA 4715
Db 130 ----- 130
QY 4714 TGAGTGGTGGGCTGGGGGCGAGGCGCTTGGGAGCTGCCA----- 4673
Db 131 ---ThrGlyAlaProGlyLeuLysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaPro 149
QY 4672 ---CCCTACACCCCAAAATGCTACTGCTATTGTTAAACTTTTCAGGAAATCTCTGTGTGGC 4616
Db 150 GlyProMetGlyProArg----- 155
QY 4615 TATGTTTGCCTCCCGAGCTGGCAACCCACAGATACCTCGGAAAGGGGCGAGAGAAGA 4556
Db 156 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg--- 173
QY 4555 GGCACCGTAATACTGGAAGACAATTCGAGGCAAGGTCCTGATGCTCAGCTCGGGTTTGT 4496
```



```
QY 2416 GCGTGGGCCCCAGGAAAGGATGAGGAGATGTATTAAGTGGGAAAGAGGAGCA 2357
      |||||:::
      ---LysGlyGlu 688
Db 680 ArgGlyAlaProGlyGlu-----
QY 2356 GGGAGAGGCTGGGATAGCAAGGCCCCAGGGGAGGTTTACACGCTGACAAAGCACAGA 2297
      |||||:::
      ---
Db 689 GlyGlyProProGlyValAlaValProProGlyGlySerGlyProAlaGly-ProProG 708
      |||||:::
      ---
QY 2296 GACAACTAATTTCTCGGTGTCTCTGGGGCTGGAGCCAGCTCCCGCCACCTCTTCTC 2237
      |||||:::
      ---
Db 708 y-----ProGlnGlyValysGlyGluArgGlySerProGlyGlyProGlyAl 724
      |||||:::
      ---
QY 2236 CTTTTTCTTCTCCGCTCCAGTAGGCTCTGGGATGTGAAGAGGAAACACGGGAAGAG 2177
      |||||:::
      ---
Db 724 aAlaGlyPhePro-----GlyAlaAr 731
      |||||:::
      ---
QY 2176 GGGG---CGGTTCCCACTGGTCTCGGCTCTCAACCTCTCTCTCTCTCTCTCTCTCT 2120
      |||||:::
      ---
Db 731 gGlyLeuProGlyProProGlySerAsnGlyAsnProGly---ProProGlyProSerG 750
      |||||:::
      ---
QY 2119 CTCTGTGAACACTCTGGACACCCCAATCTTACAGCATAACTCCATCCACACCCCT 2060
      |||||:::
      ---
Db 750 ySerProGlyLysAspGly---ProProGlyProAlaGlyAsnThrGly----- 765
      |||||:::
      ---
QY 2059 AAACCCCTGCCACTACTCGCTGGGTTCTCCTACAGTCCCTCTCCCATGACCCCTCTAG 2000
      |||||:::
      ---
Db 766 -----AlaProGlySerProGlyVal-SerGlyProLysGlyAspAlaG 780
      |||||:::
      ---
QY 1999 GAGGCCCACTAGTAGGAGGAGCAGAGGAAGGAGGCGCCCTGGTTCAGAGCAGCAG 1940
      |||||:::
      ---
Db 780 lGlnProGlyGluLysGlySerProGlyAlaGlnGlyProProGly-----AlaProG 798
      |||||:::
      ---
QY 1939 GCGCC-----CAGTCTGGGGCAGAGCGAGCTGGAACGACGCGCATGGGAGG 1892
      |||||:::
      ---
Db 798 lProLeuGlylAlaGlylIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetP 818
      |||||:::
      ---
QY 1891 TGGGGCCCCCTC---TGTCCTCCCAACCCCGCCGCGGAGAGACACCAACACACCGGTG 1835
      |||||:::
      ---
Db 818 roGlyProArgGlySerProGlyProGlnGlyValysGlyGlySerGlyLysProGlyA 838
      |||||:::
      ---
QY 1834 TAAAC-----AGAGGAGACTTCCCGGCCCCAGGGGAGGCGCGGC----- 1794
      |||||:::
      ---
Db 838 laAsnGlyLeuSerGlyGluArgGlyPro---ProGlyProGlnGlyLeuProGlyLeuA 857
      |||||:::
      ---
QY 1793 --GGCTGGGCTGCACCCAGTGGGCAATGGGCAAGCGGCTCGAGGCTGCCACAT 1736
      |||||:::
      ---
Db 857 laGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeuPro---- 875
      |||||:::
      ---
QY 1735 GCCTAGTCTGGGGCCAGGATGGGAAGTGGGGGGGCGCCACCTAAGAGGGAAGGCGCAGGA 1676
      |||||:::
      ---
Db 876 -----GlyArg---AspGlySerProGlyGlyLysGlyAspA 887
      |||||:::
      ---
QY 1675 GATAGACACAGAGGGGAGTGTGGGAGCCAAAGAGCCCTCTCACCCTCATTTTAC 1616
      |||||:::
      ---
Db 887 rg-----GlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisProG 903
      |||||:::
      ---
QY 1615 CAGAACCTCCAGGTAGGACTGTCTTCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 1556
      |||||:::
      ---
Db 903 lProProGly-----ProValGlyProA 911
      |||||:::
      ---
QY 1555 ATTCTACGCTGGATCTACATAGTGTCCAGCCAGCTCGGAGTGGTGTCTCCAGTA 1496
      |||||:::
      ---
Db 911 laGlyLysSerGlyAspArgGlyGlySerGlyProAlaGlyProAlaGlyAlaProGlyP 931
      |||||:::
      ---
QY 1495 GCGGGTGAGGTATAGTTTTCTGGATGGAGGGG-----CAGGCGCTGGGTCTCTGTGC 1440
      |||||:::
      ---
Db 931 ro-----AlaGlySerArgGlyAlaProGlyProGlnGlyProArgG 945
      |||||:::
      ---
QY 1439 GATTGAGAGCTGGCACTGCGAGGAGGTGCGAGCAGCCGTCAGCAGGCACTTAACATCC 1380
      |||||:::
      ---
Db 945 lyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlylIleLysGlyHisArgGlyPheP 965
      |||||:::
      ---
```

```
QY 1379 CC-----CACAGTCCCTGTGGGCGAGGATGGACGAGAGCATG 1341
      |||||:::
      ---
Db 965 roGlyAsnProGlyAlaProGlySerProGlyProAlaGlyGlnGlnGlyAlaIleGlys 985
      |||||:::
      ---
QY 1340 AGCGGCGCCG---GGCCACACGCGGAAGCAGCTGGCAGCCCTCAGGTCTCTCTCCCC 1284
      |||||:::
      ---
Db 985 erProGlyProAlaGlyProArgGlyProVal-----GlyProSerGlyProProGlyL 1003
      |||||:::
      ---
QY 1283 ACCTTGGCTGGGTGGGGGTGAGGGT-----GGGACACACAGATGGGAGGAGGAGGA 1230
      |||||:::
      ---
Db 1003 yAspGlyThrSer-GlyHisProGlyProIleGlyProProGlyProArgGlyAsnArg 1022
      |||||:::
      ---
QY 1229 TGAAGAGGGAAC-----ACTCTGGCAAGACGGGAGAG 1194
      |||||:::
      ---
Db 1023 GlyGluArgGlySerGlyGlySerProGlyHisProGlyGln 1036
      |||||:::
      ---
RESULT 10
US-09-570-573-21
; Sequence 21, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qiust, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-09-570-573-21
Alignment Scores:
Pred. No.: 7.69e-16 Length: 1078
Score: 330.50 Matches: 319
Percent Similarity: 29.53% Conservative: 82
Best Local Similarity: 23.49% Mismatches: 502
Query Match: 3.54% Indels: 456
```

DB: 4 Gaps: 71  
US-09-931-704-3 (1-5087) x US-09-570-573-21 (1-1078)  
QY 5059 GGAACCTGGGGTGGACAGAGCTGTGGGAGGACTGAGGGATCAAGGGTTGGTGT 5000  
Db 69 GlyProProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly 84  
QY 4999 GGCAGCAGCAGTGGGTAACTGTAGGTGGGCCCATCGCCCTGAGTGGGGGAGGAAGAGG 4940  
Db 85 GluSer-----GlyArgProGlyArgPro-----GlyAspArgGly 96  
QY 4939 ----CTCTTCCAGAGCTCTGTGCAGCTCCACAGAGGCAAGCAGGAGTG----- 4892  
Db 97 LeuProGlyProProGlyIle-----LysGlyProAlaGlyIleProGlyPhe 112  
QY 4891 ---GGAGCCAGGGTGGCAAGAGCTGTATGATCAGCACCAGGAGGGGTGAGGCTGGGCAAGA 4835  
Db 113 ProGlyMetLysGlyHisArgGly-----PheAspGlyArgAsnGlyGluysGlyGlu--- 130  
QY 4834 GACTAAGTTCTGGGCAAGCATTTTATTGTTAATAACAAGATAGAAATTTCTGCAATAAA 4775  
Db 130 ----- 130  
QY 4774 TATCATCTTAATAATAACATCTCCAATAATAATAATAATAATAACAACACTTAGAGTCA 4715  
Db 130 ----- 130  
QY 4714 TGAGTGGTGGGCTGGGGGAGGAGGCTTGGGAGCTGCCA----- 4673  
Db 131 ---ThrGlyAlaProGlyLeuLysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaPro 149  
QY 4672 ---CCCTACACCCCAAAATGCTACTGCTGTTAACTTTTCAGGAAATCTGTGTGTGGC 4616  
Db 150 GlyProMetGlyProArg----- 155  
QY 4615 TATGTTGCTCCCGAGCTGGCAACCCACAGATACCTCTGGGAAAGGGGAGAGAGA 4556  
Db 156 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg--- 173  
QY 4555 GGCACCGTAATACTGAAACAATTCGAGCAAGTCTCTGATGCTCAGCTCGGGTTTGT 4496  
Db 174 -----GlyAsnAspGlyAlaArgGlySerAspGlyGlnProGlyProPro 188  
QY 4495 TTGCCACTCTGCTTTGGAGACTCTCTGGGTCTGTGTCTGCTCTCCCTGCCCTGTGCT 4436  
Db 189 GlyProProGlyThrAlaGlyPheProGly----- 198  
QY 4435 TTCAGTATTTCTCTCTCTGTTATAGGCACCTCCCAATGCAGACCCCAACCTGCACTG 4376  
Db 199 -----SerProGlyAlaLysGlyGluValGlyProAlaGlySerPro----- 212  
QY 4375 TTAATCTCATCTCTTTTGTGTAGAATTGCCACCATGTTTGTCTTCTGAATGTGATGAT 4316  
Db 212 ----- 212  
QY 4315 AGAGTGATTGGGATAAGGAGCAACACCCATCCAGAGAGAGTGGGGGAGCCCAAGAGGC 4256  
Db 213 -----GlySerAsnGlyAlaProGlyGlnArgGlyGluPro----- 224  
QY 4255 TTCACCCCAAGAGGATGTGTGGTGGCCCTCCAAACCTGCCCTCCAGATGTGCCA 4196  
Db 225 ---GlyProGlnGlyHisAlaGly-----AlaGlnGlyProProGly----- 237  
QY 4195 CCTGAGGGCTGGGTGGGCACTGCCAGTGTAGGCAAGAGCAGAGGCGCATGAGCGC 4136  
Db 238 ProProGlyIleAsnGlySer---ProGlyGlyLysGlyGluMetGly----- 252  
QY 4135 CTCTGCACCAACCTGAACCACTTCACACTCCCTCGAGCAGTCTCTCTGTAGAAACAGGA 4076  
Db 253 -----ProAlaGlyIleProGlyAla-----ProGlyLeuMetGly 264  
QY 4075 CAGGACAGGGCCTACTGTACTCTCTCCAGCTCGGTAGACCT-----TTGGGAGGTGG 4022

Db 265 -AlaArgGlyProProGlyProAlaGlyAlaAsnGlyAlaProGlyLeuArgGlyGlyAl 284  
QY 4021 GGAGAGCAGAGCGGTGATCGCATCACAGCCCGGTCCAGGAAAGGGCCAGAGGCTC 3962  
Db 284 aGlyGluProGlyLysAsnGlyAlaLysGlyGluProGlyProArgGlyGluArgGlyG 304  
QY 3961 ACAGCTTCTGTCT-----CTGGCTCAACAGGTGTT---GGCATACAGGGCTGGCTCTC 3911  
Db 304 uAlaGlyIleProGlyValProGlyAlaLysGlyGluAspGlyLysAspGly-SerProG 324  
QY 3910 ACAAGTGGGAGCAGGTTTGAAGGGGAGCAAGAGAGAGTCCAGAGTCAGAAAGCC 3851  
Db 324 lyAspProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyAlaLeuGlySerA 344  
QY 3850 ATGAGCCCCCAGGTGCGAGGTGACTGCGAGCTCTGGAGGCTGCATCTTCTTGTAGCCG 3791  
Db 344 rgGlyProAlaGly----- 348  
QY 3790 GTTGAAGTCTTGGCCGAGCCACAGCAGGTCTGCAGTCTCTTTCAGCAGCCAGAGTC 3731  
Db 349 -----ProAsnGlyIleProGlyGluLysGlyProAlaGlyGluArgGlyA 364  
QY 3730 GTCCATCTTCTGGAGGAAGTCACTGTGGGAGGCGCAGGAGTCCAGGTGGGTTCAGTCCC 3671  
Db 364 laProGlyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArg----- 379  
QY 3670 AGGAGCGGCTGGGCGAGTGTAGCCAGAGCTCCCATGACCGCCCAATGTGCCCCAG 3611  
Db 380 -----AspGlyValProGlyGlyProGlyMetArgGlyMetProGly---SerProG 396  
QY 3610 CAGCCCTGAGGCTGTGTGCAGAAAGTGGGCCAGGCTGCGGCGCAGCTCAGCAGTGGCAGC 3551  
Db 396 lyGlyProGlySerAspGlyLysProGlyPro----- 406  
QY 3550 CTGAGGTTGAGGCCACCGCAAGTAACAGAAAGTGGGTGTAGGCTCTAGTCTGGGT 3491  
Db 407 -----ProGlySerGlnGly 412  
QY 3490 CAGCGCAGTTTGTTCATTGAGGCTTTCGCACACCTCCAAAGTCAACAGTGGCCCTGGGCG 3431  
Db 412 luSer-----GlyArgProGlyProProGlyProSerGlyProArgGlyG 427  
QY 3430 AGTCTGCCCCCAGCGGGGAGGTTGAAGTCTGGCTCGTTGAAAGGGGGGCCAGGTA 3371  
Db 427 ln-----ProGlyValMetGly-----PheProGlyProLysGlyAsnAspGlyA 442  
QY 3370 GTTCACTGTGAAAGAGAGAGGTTGATGGGGAAGAGAGGAGGCGCAGCGCTGGCTCACC 3311  
Db 442 laProGlyLys-AsnGlyGluArgGlyProGlyGly----- 454  
QY 3310 ACCAAGGAGATACCTGCTCCCAAGTTTCTATTTTGTGTCCTCCCTGACACATGGCCCTGT 3251  
Db 455 -----ProGlyProGlnGlyPro-----ProGly 462  
QY 3250 CTCATGCTAGGCTCTTTTGTTCATGGCTCCCTCGATGTTTTCCTAATAAGGATATCATTT 3191  
Db 463 LysAsnGlyGluTyrGlyProGlnGlyProPro----- 473  
QY 3190 GTATGAGGTGACATGTTGAAGCAGTGGCGGTGGTGAGCTCTCCAATGGAGAGCGGTAGAA 3131  
Db 474 -----GlyPro 475  
QY 3130 AAAGGAGGAGGAGAGCGCACAGAC---GCTCAGAAAGGAGCTGCTGTTGGCTGGGT 3074  
Db 476 ThrGlyProGlyGlyAspLysGlyAspThrGlyProArgGlyProGlnGlyLeuGlnGly 495  
QY 3073 GGGCAACCTTTGGGAGAGTGGCCGCTGTGGGGGAGGAGAGAGAGAGAGAGAGCTGGA 3014  
Db 496 LeuProGlyThrGlyGlyProPro-----GlyGluAsnGluLysProGlyGluProGly 513  
QY 3013 GGAGCTGGAGAGGTAGAAAGGTGGCCAGAGAGTGCAGAGAGTGCAGAGAACAGAGAGGTGCA 2954

Db 514 ProLeuGlyGluAla---GlyAlaProGlyAlaProGlyGlyLysGlyAspAlaGlyAla 532  
Qy 2953 GAAGGAGCAGCGCTCAGCAGTGTGGCTGCAAAACACAGACGACGAGGAGATGACCTTCA 2894  
Db 533 ProGly-----GluArgGlyProPro 539  
Qy 2893 AAACACTGTGGGCT-----TGGCAAGAGGTGCTTACGACCTTGGAGAACCACTTG 2840  
Db 540 GlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyPro----- 554  
Qy 2839 CTGGGCGCTGGAAGCCAGATGTCAGGGGGTTAAGAAGTGTGGTGGAGGAGGAGTGA 2780  
Db 555 ---GlyPro-----GluGlyGlyLysGlyAlaAlaGlyProProGlyProPro 569  
Qy 2779 GGCAGAGGTGTAGACACTCTTGGAGAAATTGGCAGGAGGAAAGCCAGGAGAGAAA--- 2723  
Db 570 GlyAlaAlaGlyThrProGlyLeu-----GlnGlyMetProGlyGluArgGly 585  
Qy 2722 -----AAGGGTTCGGAACTAGAGAGGGCAGTGGAGTGAAGCAAA 2694  
Db 586 GlyLeuGlySerProGlyProLysGlyAspLysGlyGluProGlyGlyProGlyAlaAsp 605  
Qy 2683 GGTTCCTGCACAGAAACAGAGAGATGAGAGTGGCGCCATTGGAACACAGAGCCCAT 2624  
Db 606 GlyValProGlyLysAspGlyProArgGlyProThrGlyProLleGlyProProGly--- 624  
Qy 2623 TTGCGCTGACACACAGCCAGCCACAGCTGAGGGCTGGGAGAGACGGGTACTACTA 2564  
Db 625 ---ProAlaGlyGlnProGlyAspLysGlyGluGly---GlyAlaProGlyLeuProGly 642  
Qy 2563 ACACCTCCGAAGAGGGTTCA---GCTGATAGGAGCAGGGTT----- 2525  
Db 643 IleAlaGlyProArgGlySerProGlyGluArgGlyGluThrGlyProProGlyProAla 662  
Qy 2524 -----CGAAGAGAAATGAGAGGAGTCTCATGATGAGACAGAGGAGGAGG 2477  
Db 663 GlyPheProGlyAlaProGlyGlnAsnGly-----GluProGlyGly--- 676  
Qy 2476 CTATGAAGACACATGGTATAGGAAGAGGCGAGAGCCCTTTCCACCCCTCGGGGAGC 2417  
Db 677 -----LysGlyGlu 679  
Qy 2416 GGCTGGGCGCCAGGAAAGGATGAGGAGATGTATACTGGGAAAGAGGAGCA 2357  
Db 680 ArgGlyAlaProGlyGlu-----LysGlyGlu 688  
Qy 2356 GGGGAGGCGCTGGATAGCAGAGGCGCCAGGGGAGGTTCACAGCTGACACACAGA 2297  
Db 689 GlyGlyProProGlyValAlaValProProGlyLysSerGlyProAlaGly-ProProGl 708  
Qy 2296 GACAACTAATTTCTCGGTGTATCTCGGGCTGGAGCCAGCTCCCGCCACCTTCTTC 2237  
Db 708 y-----ProGlnGlyValLysGlyGluArgGlySerProGlyGlyProGlyVal 724  
Qy 2236 CTTTCTTCTTCCTGCTCCAGAGTAGGTCTGGGATGTGGAAGAACACACGGGAAGAG 2177  
Db 724 aAlaGlyPhePro-----GlyAlaAr 731  
Qy 2176 GGGG---CCGGTTCCTCATGTCCTCGGCTCCTCCAAACCTCTTCTCTCTCTCTCGG 2120  
Db 731 gGlyLeuProGlyProProGlySerAsnGlyAsnProGly---ProProGlyProSerGl 750  
Qy 2119 CTCTCTGAACACTCTGACACACCCCAATCTTACAGCATACCTCCATCCACACCT 2060  
Db 750 ySerProGlyLysAspGly---ProProGlyProAlaGlyAsnThrGly----- 765  
Qy 2059 AAACCCCTGCCACTACTGCTGGTTCCTACAGTCCCTTCCCTCCATGACCCCTTAG 2000  
Db 766 -----AlaProGlySerProGlyVal-SerGlyProLysGlyAspAlaG 780  
Qy 1999 GAGGCCAGCTAGATGGGAGAGCAGACGAGAGGGAGGCGCTGGTCCAGCAGCAG 1940  
Db 780 yGlnProGlyGluLysGlySerProGlyAlaGlnGlyProProGly-----AlaProG 798

Qy 1939 GGGC-----CAGGTCTGGGGCAGAGCGAGCTGGAACCGCAGGCATGGGAGG 1892  
Db 798 lyProLeuGlyLleAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetP 818  
Qy 1891 TGGGGCGCCCTC---TGTCCCAACCAACCCCGCCGCGAGAGACACACACACACGCTG 1835  
Db 818 roGlyProArgGlySerProGlyProGlnGlyValLysGlyLysGlyLysProGlyA 838  
Qy 1834 TAAAC-----AGAGGAGACTTCCCGGCCCCCAGGGGGAGGGCGGCG--- 1794  
Db 838 laAsnGlyLeuSerGlyGluArgGlyPro---ProGlyProGlnGlyLeuProGlyLeuA 857  
Qy 1793 ---GGCTGGCTGGCAGCAGCTGGGGCAATGGGCCAAGCGGCTGCGAGGCGCTGCCACAT 1736  
Db 857 laGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeuPro--- 875  
Qy 1735 GCCTAGTCTCTGGGCGCAGGATGGGAAAGTGGGGGCCACCTAAGAGGAGGAGGAGG 1676  
Db 876 -----GlyArg---AspGlySerProGlyGlyLysGlyAspA 887  
Qy 1675 GATAGACCAGACAGGGGACTGGTGGGAGCCAAAGAGCCCTCTCCACCCCATTTAC 1616  
Db 887 rg-----GlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisProG 903  
Qy 1615 CAGAACCTCTCCACGGTTAGGACTGTCTTCCCAACTCTCTACTCTCTCCCAACTCCAG 1556  
Db 903 lyProProGly-----ProValGlyProA 911  
Qy 1555 ATTCCTACGCTGGATCTCATATAGTCCAGCCCAAGCTGCGGAGTTGTGTCTCCAGGTA 1496  
Db 911 laGlyLysSerGlyAspArgGlyGluSerGlyProAlaGlyProAlaGlyAlaProGlyP 931  
Qy 1495 GCGGTGAGGTCTAGTTTCTTCTGATGGAGGG---CCAGGCGCTGGGTCCCTGTGC 1440  
Db 931 ro-----AlaGlySerArgGlyAlaProGlyProGlnGlyProArgG 945  
Qy 1439 GATTGAGAGCTGGCACTCAGGGAGTCCAGGAGCCAGTGCACAGCCGCTTAACATCC 1380  
Db 945 lyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlyLysGlyHisArgGlyPheP 965  
Qy 1379 CC-----CACGATCCCTGTGGCAGGATGGACGAGAGCATG 1341  
Db 965 roGlyAsnProGlyAlaProGlySerProGlyProAlaGlyGlnGlnGlyAlaIleGlyS 985  
Qy 1340 AGCGCGGCGCG---GGCCACACAGGGAAGCAGCTGGCAGCCCTCAGGTGTCTGTCCCC 1284  
Db 985 erProGlyProAlaGlyProArgGlyProVal-----GlyProSerGlyProProGlyL 1003  
Qy 1283 ACCTTGGCTGGTGGGGGTGAGGGT-----GGGACACACAGATGGGAGGAGGGA 1230  
Db 1003 ysAspGlyThrSer-GlyHisProGlyProLleGlyProProGlyProArgGlyAsnArg 1022  
Qy 1229 TGGAGGAGGGAAAC-----ACTCTGGCAAGACGGGAGAG 1194  
Db 1023 GlyGluArgGlySerGluGlySerProGlyHisProGlyGln 1036

## RESULT 11

US-09-548-608-21

; Sequence 21, Application US/09548608

; Patent No. 6355442

; GENERAL INFORMATION:

; APPLICANT: Qvist, Per

; APPLICANT: Bonde, Martin

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

; TITLE OF INVENTION: Disorders Associated with the Metabolism of

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby &amp; Darby PC

; STREET: 805 Third Avenue

; CITY: New York

STATE: New York  
 COUNTRY: USA  
 ZIP: 10022  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/548,608  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/187,319  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gogoris, Adda C  
 REGISTRATION NUMBER: 29,714  
 REFERENCE/DOCKET NUMBER: 4305/08701  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-527-7700  
 TELEFAX: 212-753-6237  
 TELEX: 236687  
 INFORMATION FOR SEQ ID NO: 21:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1078 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: COLLAGEN ALPHA 1 (III)  
 US-09-548-608-21

Alignment Scores:  
 Pred. No.: 7,69e-16 Length: 1078  
 Score: 330.50 Matches: 319  
 Percent Similarity: 29.53% Conservative: 82  
 Best Local Similarity: 23.49% Mismatches: 502  
 Query Match: 3.54% Indels: 456  
 DB: 4 Gaps: 71

US-09-931-704-3 (1-5087) x US-09-548-608-21 (1-1078)

QY 5059 GGGAACTGGGTGGACAGAGCTGTGGGAGGACTGAGGATCAAGGGTTGGTGT 5000  
 Db  
 69 GlyProProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly 84  
 QY 4999 GCGACGAGCTGGGTAACTCTGAGTGGCGCCATGCCCGGAGGAGGAGG 4940  
 Db  
 85 GluSer-----GlyArgProGlyArgPro-----GlyAspArgGly 96  
 QY 4939 ---CCTCTTCCAGAGCTCTGTGCGAGCTCCACAGAGGCAAGCAGGAGTG- 4892  
 Db  
 97 LeuProGlyProProGlyle-----LysGlyProAlaGlyIleProGlyPhe 112  
 QY 4891 ---GGAGCCAGGGTGGCAGAGCTCTGATGAGCAGCAGGAGGGGTGAGGCTGGGCAAGA 4835  
 Db  
 113 ProGlyMetLysGlyHisArgGly---PheAspGlyArgAsnGlyGluLysGlyGlu- 130  
 QY 4834 GACTAAGTTCTGGGCAAGCATTTATTGTTTAAATACAGAAATAGAAATCTGCAATAAA 4775  
 Db  
 130 ----- 130  
 QY 4774 TATCATCTAATAATAACATCTCCAAATAATAATAATAATAACACACTTAGAGTCA 4715  
 Db  
 130 ----- 130  
 QY 4714 TGAGTGGGTGGGCTGGGGGCGAGGCGCTTGGGAGCTGCCA- 4673  
 Db  
 131 ---ThrGlyAlaProGlyLeuLysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaPro 149

QY 4672 ---CCCTACACCCCAAAATGCTACTGCAATTGTAAACTTTTTCAGGAAATCCTGTGTGTGGC 4616  
 Db  
 150 GlyProMetGlyProArg----- 155  
 QY 4615 TATGTTGCCCTCCCGAGCCTGGCAACCCACAGATACCTCTGGGAAAGGGGCGAGAGAAGA 4556  
 Db  
 156 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg- 173  
 QY 4555 GGCACCGTAATACTGGAACACAAATCGAGGCAAGGCTCTGATGCTGAGCTCGGTTTGT 4496  
 Db  
 174 -----GlyAsnAspGlyAlaAlaArgGlySerAspGlyGlnProGlyProPro 188  
 QY 4495 TTGCCACTCTGTGCTTTGGAGACTCCTGGTCTGGTCTGTCTGTCTCCTCCCTGCTGCT 4436  
 Db  
 189 GlyProProGlyThrAlaGlyPheProGly----- 198  
 QY 4435 TTCAGTATTTCTTCTGTTTATAGGCACCTCCAAATGCAGACCCCAACCTTGCACCTG 4376  
 Db  
 199 -----SerProGlyAlaLysGlyGluValGlyProAlaGlySerPro----- 212  
 QY 4375 TTAATCTCATCTCTTTTGTGTAGATTGCCACCATGTTTGTCTTCTGAATTGATGTAT 4316  
 Db  
 212 ----- 212  
 QY 4315 AGAGTGATTGGGGATAAGGGAGCAACACCCATCCAAAGAGAGTGGGGCAGCAAGAGGC 4256  
 Db  
 213 -----GlySerAsnGlyAlaProGlyGlnArgGlyGluPro----- 224  
 QY 4255 TTCACCCAGAAAGCATGTGTGGTGGCCCTCAACCCCTGCCCTCCAGATGTGCCA 4196  
 Db  
 225 ---GlyProGlnGlyHisAlaGly-----AlaGlnGlyProProGly----- 237  
 QY 4195 CCTGAGGGCTGGGTCGTCAGTGGCCACTGGCCNAGTGTAGGCAGAAAGCAGGAGCCATGACGC 4136  
 Db  
 238 ProProGlyIleAsnGlySer---ProGlyGlyLysGlyGluMetGly----- 252  
 QY 4135 CTCTGCACCAACTGAACCATTTCACTCTCCCTCCAGCATGACTTCTGTAGAAACAGGA 4076  
 Db  
 253 -----ProAlaGlyIleProGlyAla-----ProGlyLeuMetGly 264  
 QY 4075 CAGGACAGGGCTACTGTACTCTCTCCCGAGCTCGGTAGACCT-----TTGGGAGGTGG 4022  
 Db  
 265 -AlaArgGlyProProGlyProAlaGlyAlaAsnGlyAlaProGlyLeuArgGlyGlyAl 284  
 QY 4021 GGAGGACAGCGGTGATCGCATCACACCCCGCGGTCCAGGAAAGGGCAGAGGCTC 3962  
 Db  
 284 aGlyGluProGlyLysAsnGlyAlaLysGlyGluProGlyProArgGlyGluArgGlyGly 304  
 QY 3961 ACAGTTCTGTCT-----CCTGGCTCAACAGGTGT---GGCATACAGGGCTGGCTCTC 3911  
 Db  
 304 uAlaGlyIleProGlyValProGlyAlaLysGlyGluAspGlyLysAspGly-SerProG 324  
 QY 3910 ACAAGTGGGAGCGGGTTTGAAGGGGAGCGAGAGGAGGTCAGAAAGTCAGAAAGCC 3851  
 Db  
 324 LysAspProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyAlaLeuGlySerA 344  
 QY 3850 ATGAGCCCCAGGTGCGAGGTGACTGTCAGCTGCTGGAGGCTGCATCTTCTTGTAGCCG 3791  
 Db  
 344 rGlyProAlaGly----- 348  
 QY 3790 GTTGAAGTCTTGGCCGAGCGCCACAGCCAGGTCTGAGCTCTTTCAGCAGCCAGAAAGTC 3731  
 Db  
 349 -----ProAsnGlyIleProGlyGluLysGlyProAlaGlyGluArgGlyAla 364  
 QY 3730 GTCCATCTTCTGGAGGAGTCACTGTGGGAGGGCCAGGAGTCCAAAGTGGGTTCAGTCCC 3671  
 Db  
 364 laProGlyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArg----- 379  
 QY 3670 AGGCAGCGGCTGGGCGAGGTAGCCAGAGCTGCCATGACCCCGCAATCTGCCAG 3611  
 Db  
 380 -----AspGlyValProGlyGlyProGlyMetArgGlyMetProGly---SerProG 396  
 QY 3610 CAGGCCCTGGAGGCTGTGTGCAAGAGTGGGCCAGGCTGGGGCGAGCTCAGCAGTGGCAGC 3551

Db	396	lyGlyProGlySerAspGlyLysProGlyPro	406
QY	3550	CTCAGCGTTGAGGCCACGCAAGTAACACAGAGGTGCTGTAGCGCTCGTAGTCTGGGT	3491
Db	407	ProGlySerGlnGlyG	412
QY	3490	CAGCCGCGAGTTTGTCATTGAGGCTTCGCCACACCTCCAAGTCAACAGTGGCCCTGGGCAG	3431
Db	412	luser	427
QY	3430	AGTCTCTGCCCCAGCGGGGAGGGGTTGAAGTCTGGCTCGTTGAAAGGGGGGCCAGGTA	3371
Db	427	ln	442
QY	3370	GTTCCAGCTGTAAAGAGAGAGGGTGATGGGAAAGGAGAGGCGAGCGGCTGGCTCACC	3311
Db	442	laProGlyLys-AsnGlyGluArgGlyProGlyGly	454
QY	3310	ACCAAGGAGATACCTGCTCCCAAGTTTCTATTTTGTGTCCCTGACACTGGCCCTGT	3251
Db	455	ProGlyProGlnGlyPro	462
QY	3250	CTCCATGCTAGGCTTCTTTGTTTCATGGCTCCCTGAGTTTTTCCAATAAGGATATCATTT	3191
Db	463	LysAsnGlyGluTyrGlyProGlnGlyProPro	473
QY	3190	GTATGCAAGTGACAGTTGAAGCAGTGGCGGTGTGAGCTCTCCAATGAGAGCGGTAGAA	3131
Db	474	GlyPro	475
QY	3130	AAAGGAGGAGGAGAGCAGCAAGAC--GCTCAGGAAGACGCTGCTGATTGGCTGGGT	3074
Db	476	ThrGlyProGlyGlyAspLysGlyAspThrGlyProArgGlyProGlnGlyLeuGlnGly	495
QY	3073	GGGCAAACTTGGGAGAGTGGCCGCTGTGGGGGACGACAGAGAGGAAGAGAGCTGGA	3014
Db	496	LeuProGlyThrGlyGlyProPro--GlyGluAsnGluLysProGlyGluProGly	513
QY	3013	GGAGCTGGAGAGGTAGAAGTGGCCAGGAGAGTGCAGAGACCAAGGAGAGAGGTTCCA	2954
Db	514	ProLysGlyGluAla--GlyAlaProGlyAlaProGlyGlyLysGlyAspAlaGlyAla	532
QY	2953	GAAGCAGCAGCGCTCAGCAGTGTGGCTGCTCAAAACAGACAGCAAGGAGAATAGCCTTCA	2894
Db	533	ProGly--GluArgGlyProPro	539
QY	2893	AAACACTGTGGGCT--TGGCAAGAGGTGCTGACACTTTGGAGAGAACATTG	2840
Db	540	GlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyProPro	554
QY	2839	CTGGGGCTGGAAGCCAGCATTCACGGGGTTAAGAGTCAGTGGTGGAGAGAACTGGA	2780
Db	555	GlyPro--GluGlyGlyLysGlyAlaAlaGlyProGlyProProPro	569
QY	2779	GGCAGAGGGTGTAGACCACTCTCTTGGAGAAATTTGGCAGGAAAGCCAGAGAGAAA--	2723
Db	570	GlyAlaAlaGlyThrProGlyLeu--GlnGlyMetProGlyGluArgGly	585
QY	2722	--AAGGTCGGGAACCTAGAGAGGGCAGTGGAGTGAAGCAAA	2684
Db	586	GlyLeuGlySerProGlyProLysGlyAspLysGlyGluProGlyGlyProGlyAlaAsp	605
QY	2683	GTTTTCTCCACAAAGAAACAGAGAGATGACAGTGGCCCATTTGGNAACAGAGCCCCAT	2624
Db	606	GlyValProGlyLysAspGlyProArgGlyProThrGlyProIleGlyProProGly--	624
QY	2623	TTGGCGCTGACACACCCAGGCCCCACAGCTGAGGGGCTGGGGAGAGACGGGTACTCACTA	2564
Db	625	ProAlaGlyGlnProGlyAspLysGlyGluGly--GlyAlaProGlyLeuProGly	642
QY	2563	ACACTCCGAAGAGGTTTCA--GCTGTATGAGCAGCGGTT--	2525

Db	643	IleAlaGlyProArgGlySerProGlyGluArgGlyGluThrGlyProProGlyProAla	662
Qy	2524	-----GCAGAAGGAAGAAATGGAGCGACTCTCATGAGTAGGACAGGAGGAAGG	2477
Db	663	GlyPheProGlyAlaProGlyGluAsnGly	676
Qy	2476	CTATGGAAAGACAGCATGTTATAGGAGAGGGCAGAGCCCTTCCAAACCTCGGGGAGC	2417
Db	677	-----LyGlyGlu	679
Qy	2416	GGCTGGGCCCGAGGAAAGGATGAGGAGATGTATACTGGGAAGAGGAGAGAACAA	2357
Db	680	ArgGlyAlaProGlyGlu	688
Qy	2356	GGGAGAGCGCTGGGATAGCAGAGGGCCAGGGGAGGTTCACACGCTCACACAGCACAGA	2297
Db	689	GlyGlyProProGlyValAlaValProProGlyGlySerGlyProAlaGlyProProGly	708
Qy	2296	GACAACTAAATTTCTCGGTGTATCTCGGGCTGGAGCAGCTCCCGCCACCTCTTCT	2237
Db	708	y-----ProGlnGlyValLyGlyGluArgGlySerProGlyGlyProGlyAl	724
Qy	2236	CTTTTCTTCTCCCTGCCTCCAGAGTAGGCTCGGGATGTGGAAAGAGAACACGGGAAG	2177
Db	724	aAlaGlyPhePro	731
Qy	2176	GGGG---CGGTTTCCCATGGTGCCTCGGCTCTCCAAACCTCTTCTCTCTCTCTCGG	2120
Db	731	gGlyLeuProGlyProProGlySerAsnGlyAsnProGly	750
Qy	2119	CTCTCTGAACACTCTCGGACCAACCCCAATCTTTACAGCTAACTCCATCCACACCCT	2060
Db	750	ySerProGlyLyAspGly---ProProGlyProAlaGlyAsnThrGly	765
Qy	2059	AAACCCCTGCCACTACTCGCTGGGTTCCTACAGTCCCTTCCCCCATGACCCCTCTAG	2000
Db	766	-----AlaProGlySerProGlyVal-SerGlyProLyGlyAspAlaG	780
Qy	1999	GAGGCCACGCTAGGTGGAGAGGACAGCGAAGAGGAGGCCCTCGGTCCAGAGCAGCAG	1940
Db	780	LyGlnProGlyGluLyGlySerProGlyAlaGlnGlyProProGly	798
Qy	1939	GGCCC-----CAGTCTGGGGCAGAGGCGGCTGGAACGCAGCATGGGAGG	1892
Db	798	LyProLeuGlyIleAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetP	818
Qy	1891	TGGGGCCCTC--TGTCGCCACCAACCCGCCCGCGCAAGAGACACACACACGCGTG	1835
Db	818	roGlyProArgGlySerProGlyProGlnGlyValLyGlyGlyGlySerGlyLyAspGly	838
Qy	1834	TAAAC-----AGAGGAGACTTCCCGGCCCCCGAGGGGAGCGCGGC-----	1794
Db	838	LaAsnGlyLeuSerGlyGluArgGlyPro---ProGlyProGlnGlyLeuProGlyLeuA	857
Qy	1793	---GGCTGGGTGCACCCAGTGGGGCAATGGGCCAAGCGGGTGCAGAGGCTGCCACAT	1736
Db	857	laGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeuPro---	875
Qy	1735	GCCTAGTCTTGGGGCCAGGATGGGAAGTGGGGGGGCCACCTTAAGAGGGAAGGCGCAGGA	1676
Db	876	-----GlyArg---AspGlySerProGlyGlyLyGlyAspA	887
Qy	1675	GATAGACCACAGGGGAGCTGTGGGAGCAAGAGCCCTCTCACCCCATCATTTAC	1616
Db	887	rg-----GlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisProG	903
Qy	1615	CAGAACCTCCACGGTTAGGACTGTCTTCCCAACTCTCTCACTCTCTCCCACTCCAC	1556
Db	903	LyProProGly-----ProValGlyProA	911
Qy	1555	ATTCTACGCTGGATCTACATAGTTCACAGCCCAAGCTGGGAGTTGGTCTCCAGTA	1496
Db	911	laGlyLyAspGlyAspArgGlyGluSerGlyProAlaGlyProAlaGlyAspProGlyP	931



QY 1495 GCGGTCAGGTCATAGGTTTCTGTGATGAGGGG---CCAGGCCCTGGGTCCCTGTGCT 1440  
 Db 931 ro-----AlaGlySerArgGlyAlaProGlyProGlnGlyProArgg 945  
 QY 1439 GATTGAGAGCTGGCAGCTGCGAGGAGTCCAGAGCACCGTGCACAGGACGCTAACATCC 1380  
 Db 945 lyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlyIleLysGlyHisArgGlyPheP 965  
 QY 1379 CC-----CACGAGTCCCTGTGGGAGGATGACGAGAGCATG 1341  
 Db 965 roGlyAsnProGlyAlaProGlySerProGlyProAlaGlyGlnGlnGlyAlaIleGlys 985  
 QY 1340 AGCCGCGCCCG---GGCCCCACACGGGGAAGCAGCTGGCAGCGCCCTCAGGTGCTGTCCCC 1284  
 Db 985 exProGlyProAlaGlyProArgGlyProVal-----GlyProSerGlyProProGlyL 1003  
 QY 1283 ACCTTGGCTGGTGGGGGTGAGGT-----GGGACACAGATGGGAGGAGGAGGGA 1230  
 Db 1003 ysAspGlyThrSer-GlyHisProGlyProIleGlyProGlyProArgGlyAsnArg 1022  
 QY 1229 TGGAGGAGGGGAAC-----ACTCTGGCGCAAGCGGAGAG 1194  
 Db 1023 GlyGluArgGlySerGluGlySerProGlyHisProGlyGln 1036

## RESULT 12

US-08-931-820-4  
 ; Sequence 4, Application US/08931820  
 ; Patent No. 6010863  
 ; GENERAL INFORMATION:  
 ; APPLICANT:  
 ; TITLE OF INVENTION: Assay for collagen degradation  
 ; NUMBER OF SEQUENCES: 4  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/931,820  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: EP 96202596.1  
 ; FILING DATE:  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1057 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo sapiens  
 ; TISSUE TYPE: Collagen type III  
 ; FEATURE:  
 ; NAME/KEY: Modified-site  
 ; LOCATION: 1055  
 ; OTHER INFORMATION: /label= Modified  
 ; OTHER INFORMATION: /note= "Ala may be Pro"

US-08-931-820-4

## Alignment Scores:

Pred. No.:	1,26e-15	Length:	1057
Score:	327.50	Matches:	326
Percent Similarity:	29.96%	Conservative:	80
Best Local Similarity:	24.06%	Mismatches:	501
Query Match:	3.51%	Indels:	450
DB:	3	Gaps:	71

US-09-931-704-3 (1-5087) x US-08-931-820-4 (1-1057)

QY 5059 GGGAAACCTGGGTGGACAGAACTGTGTGGAGGACTGAGGGATCAAGGGTTGGTGT 5000  
 Db 68 GlyProProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly 83  
 QY 4999 GGCAGCAGCAGTGGTAACTGTGAGGTGCGCCATCCGCCAGTGCAGGAGGAGGAGG 4940  
 Db 84 GluSer-----GlyArgProGlyArgPro-----GlyGluArgGly 95  
 QY 4939 ---CCTCTTCCAGAGCTGTGTGAGCTCCACAGAGGCAAGCAGGAGTG----- 4892  
 Db 96 LeuProGlyProProGlyIle-----LysGlyProAlaGlyIleProGlyPhe 111  
 QY 4891 ---GGAGCCAGGGTGCAGAGAGTCTGATGAGCAGCAGGAGGAGTGGTGGGCAAGA 4835  
 Db 112 ProGlyMetLysGlyHisArgGly---PheAspGlyArgAsnGlyGluLysGlyGlu--- 129  
 QY 4834 GACTAAGTTCTGGGGCAAGCATTTTATTGTTTAATACAGAAATAGAAATCTGCAATAA 4775  
 Db 129 ----- 129  
 QY 4774 TATCATCTAATAAATAACATCTCCAAATAAATAAATAAATAAATAAATAAATAA 4715  
 Db 129 ----- 129  
 QY 4714 TCAGTGGGTGGGCTGGGGGCGAGGCGCTTGGGAGCTGCCA----- 4673  
 Db 130 ---ThrGlyAlaProGlyLeuLysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaPro 148  
 QY 4672 ---CCCTACACCCCAAAATGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4616  
 Db 149 GlyProMetGlyProArg----- 154  
 QY 4615 TATGTTGGCTTCCCGCAGCCTGGCAACCCACAGATACCTTGGGAAAGGGGCGAGAGAAGA 4556  
 Db 155 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg--- 172  
 QY 4555 GGCACCGTAATACTCGAGCAACAATCGAGGCAAGTCTGATGCTCAGCTCGGTTTGT 4496  
 Db 173 -----GlyAsnAspGlyAlaArgGlySerAspGlyGlnProGlyProPro 187  
 QY 4495 TTGCCACTGTGTGTTGGAGACTCTGGGTCTGTGCTGTGCTGCTGCTGCTGCTGCTGCT 4436  
 Db 188 GlyProProGlyThrAlaGlyPheProGly----- 197  
 QY 4435 TTCAGTATTTCTCTTCTGTTTATAGGCGACCTCCAATGCAGACACCCCAACCTTGCAC 4376  
 Db 198 -----SerProGlyAlaLysGlyGluValGlyProAlaGlySerPro----- 211  
 QY 4375 TTAATCTCATCTCTTTTGTGTAGAATTGCCACCATGTTTGTCTGTAATGATGTAT 4316  
 Db 211 ----- 211  
 QY 4315 AGAGTATTGGGATAGGGAGCAACACCCATCCAAGAGAGTGGGCGAGCCAAAGGCG 4256  
 Db 212 -----GlySerAsnGlyAlaProGlyGlnArgGlyGluPro----- 223  
 QY 4255 TTCACCCCAAGAGCATGTGTGGTGGGCGCCCTCAACCCCTGCTCCAGATGTGCCA 4196  
 Db 224 ---GlyProGlnGlyHisAlaGly-----AlaGlnGlyProProGly----- 236  
 QY 4195 CTGAGGGGCTGGTGGGCACTGGCCAAAGTGTAGCAAGAGAGGAGGAGGAGGAGGAGG 4136  
 Db 237 ProProGlyIleAsnGlySer---ProGlyGlyLysGlyGluMetGly----- 251  
 QY 4135 CTCTGCACCAACCTGAACCACTTCACTTCCCTCGAGCATGACTTCTGTAGAAACAGGA 4076  
 Db 252 -----ProAlaGlyIleProGlyAla-----ProGlyLeuMetGly 263  
 QY 4075 CAGGACAGGGCTACTGTACTCTCTCCCGAGCTCGGTAGACCT-----TTGGGAGGTGG 4022  
 Db 264 -AlaArgGlyProProGlyProAlaGlyAlaAsnGlyAlaProGlyLeuArgGlyAla 283  
 QY 4021 GGAGGAGACAGGGTGTATCGCATCACAGCCCGCGGTCCAGGAAAGGGCCAGAGGCTC 3962

Db 283 aGlyProGlyLysAsnGlyAlaLysGlyGluProGlyProArgGlyGluArgGly-- 302  
QY 3961 ACAGCTTCTGCTCCCTGGCTCAACAGGTGTT---GCCATACAGGGTGGCTCTCAAAAG 3905  
Db 303 -----GluAlaGlyLeuProGlyValProGlyAlaLysGly-GluAspGlyLysA 319  
QY 3904 TGGAGCAGGGTTGAAGGGGAGCGAAGAGGAGAGGTTCAGAGTCAGAGCCATGAGC 3845  
Db 319 spGlySerProGlyGluProGlyAlaAsnGlyLeuProGlyAlaLysGlyGluArgGlyA 339  
QY 3844 CCCAGGT-----CGAGGTGACTGTCAGCTGCTGGAGGTGCTCATCTTCTTCT 3797  
Db 339 laProGlyPheArgGlyProAlaGly----- 347  
QY 3796 GAGCCGGTTGAAGTCTTGGCGAGCGCCACAGCCAGGTCTGAGCTCTTTCAGAGCCA 3737  
Db 348 -----ProAsnGlyLeuProGlyGluLysGlyProAlaGlyGluA 361  
QY 3736 GAAGTCCTCATCTTCTGGAGGAGTCACTGTGGGCGAGGCCAGGAGTCCAAAGTGGGTT 3677  
Db 361 rgGlyAlaProGlyProAlaGlyProArgGlyAlaLysGlyGluProGlyArg----- 378  
QY 3676 AGTCCAGGCGGCTGGGCGAGTGGTGGAGCCAGAGTGGCCATGACGCCGCAATGCT 3617  
Db 379 -----AspGlyValProGlyGlyProGlyMetArgGlyMetProGly---S 393  
QY 3616 GCCCAGAGCCCTGAGGCTGTGAGAGTGGGCGAGGCTGGCGCGAGCTCAGCAGT 3557  
Db 393 erProGlyGlyProGlySerAspGlyLysProGlyProArgGlySerGlnGlyGluSerG 413  
QY 3556 GCGAGCTGACGTTGAGGCCACGCAAGTAAACAGAGAGTGGCTGTAGGCTCGTAGT 3497  
Db 413 lyArgPro-----GlyProProGlyProSerGlyProArg----- 424  
QY 3496 CTGGGTGAGCGCGAGTTTGTTCATTGAGGCTTGGCCACACCTCCCAAGTCAACAGTGGCC 3437  
Db 425 ---GlyGlnProGlyValMet-----GlyPheProGlyPro-Lys----- 436  
QY 3436 GGGCAGAGTCTGCCCCAGGGGAGGAGGTGAAGTCTGGCTCGTTGAAGGGGGGCC 3377  
Db 437 -GlyAsnAspGlyAlaProGlyLysAsnGlyGluArgGlyGly-----ProGlyGlyPr 454  
QY 3376 CAGGTAGTTGAGCTGTGAAAAGGAGGAGGTGATGCGGAGGAGGAGGAGGCGGCTGG 3317  
Db 454 o-----GlyProGlnGlyProProGlyLysAsnGlyGluThrGlyProG 469  
QY 3316 CTCACCACCAAGGAGATACCTGCTCCCAAAGTTTCTATTTTGTGCTCCCT----- 3265  
Db 469 nGlyProPro-----GlyProThrGlyPr 477  
QY 3264 -----CACACTGCCCCGTCTCCATGCTAGGCTTCTTGTTCATGGCT 3221  
Db 477 oGlyLysAspLysGlyAspThrGlyProProGlyProGlnGly-----LeuGlnGlyLe 495  
QY 3220 CCTGAGTTTTTCCAATAAGGATATCATTTGTATGCAAGTTCACAGTTCGAGCGAGTGGCG 3161  
Db 495 uPro-----GlyThrGlyG 500  
QY 3160 TGTGTAGCTCTCCAATGAGAGGCGTA-GAAAAAGGAGGAGGAGGAGGAGCGCAAGAGCGC 3102  
Db 500 yProProGlyGluAsnGlyLysProGlyGluProGlyProLysGlyAspAla----- 517  
QY 3101 TCAGGAGGAGCTGCTGATTGGCTGGTGGCGCAACCTTGGAGAGTCCCCGCTGTGGG 3042  
Db 517 ----- 517  
QY 3041 GCAGGACAGAGGAGGAGGAGCTGGAGAGGAGTGGAGGTGAGAGTGGCCCGAGGA 2982  
Db 518 -----GlyAlaProGlyAlaProGlyGlyLysGlyAspAla---GlyAlaProGlyG 534  
QY 2981 GTGAGAGACCAAGAGAGGTTCGAGAGGAGGAGCGCTCAGCAGTGTGGCTGCA 2922

534 uArgGlyProGlyLeuAlaGlyAlaProGly----- 545  
2921 AACAGACAGCAAGGAGATGAGCCTTCAAAACACTGTGCGGCTTGGCAAGGAGGTGCTCTG 2862  
546 -----LeuArgGlyGlyAlaG 551  
2861 ACACCTTGGAGAGAACCATTTGCTGGGCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2802  
551 yProPro-----GlyPro-----GluGlyGlyLysGlyAl 561  
2801 GAGTGGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2742  
561 aAlaGlyProGlyProGlyAlaAlaGlyThrProGlyLeu-----G 577  
2741 GGGAAAGCCAGGAGAGAAA-----AAGGTCGGGAACATGA 2706  
577 nGlyMetProGlyGluArgGlyGlyLeuGlySerProGlyProLysGlyAspLysGlyG 597  
2705 GAGGCGAGTGGAGTGAAGAAAGTTTTCACAAAGAAACAGAGAGAGAGAGAGAGAGAGAG 2646  
597 uProGlyGlyProGlyAlaAspGlyValProGlyLysAspGlyProArgGlyProThrG 617  
2645 CCCATTGGAACACAGGCCCCATTGTCGCTGACACACCCAGGCCACAGCTGAGGGGCT 2586  
617 yProLysGlyProGly-----ProAlaGlyGlnProGlyAspLysGlyGlyGly-- 634  
2585 GGGGAGAGACCGGTACTCACTAACACTCCGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2526  
635 -GlyAlaProGlyLeuProGlyLysGlyAlaGlyProArgGlySerProGly---GluArgG 653  
2525 TGCAGAGGAGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2466  
653 yGluThrGlyProGlyProAlaGlyPheProGlyAlaProGlyGln-AsnGlyGluP 673  
2465 AGCATGTTATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2406  
673 roGlyGlyLysGlyGluArg-----GlyAlaProGlyGluL 685  
2405 AGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2346  
685 yGlyGlyGlyGlyProGlyValAlaGlyProGly-----GlyS 700  
2345 TGGAGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2286  
700 erGlyProAlaGlyProGly----- 707  
2285 TTCCTCGGTGATCTGCGGCTCTCCACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2226  
708 --ProGlnGlyValLysGlyGluArgGlySerProGlyGlyProGlyAlaAlaGlyPheP 727  
2225 CCTGCTCCAGAGTAGGCTCTGGGATGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2169  
727 ro-----GlyAlaArgGlyLeuProG 734  
2168 TTCCTCATGTCCTGCGCTCTCCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2109  
734 yProProGlySerAsnGlyAsnProGly---ProProGlyProSerGlySerProGlyL 753  
2108 CCTCTGGACCCACCCCAATCTCTTACAGCATTAACCTTCCATCCACACCTTAAACCCCTGCG 2049  
753 yAspGly---ProProGlyProAlaGlyAsnThrGly----- 764  
2048 CACTACTGCTGGTTCCTTACAGTCCCTTCCCTCATGACCCCTTCCCTCATGAGGAGGAGGAG 1989  
765 ----AlaProGlySerProGlyVal-SerGlyProLysGlyAspAlaGlyGlnProGly 782  
1988 AGGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1935  
783 GluLysGlySerProGlyAlaGlnGlyProProGly-----AlaProGlyProLeuGly 800  
1934 -----CAGTCTGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1881  
801 IleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyProProGlyMetProGlyProArg 820

QY 1880 ---TGTCCTCCACCCCAACCCCGCCGCAAGAGACACACACACACCGGTAAAC----- 1830  
Db 821 GlySerProGlyProGlnGlyValLysGlyGlySerGlyLysProGlyAlaAsnGlyLeu 840  
QY 1829 -----AGAGAGACTTCCCGCGCCCGCCAGGGAGGGGGC-----GGCTGGGCT 1785  
Db 841 SerGlyGluArgGlyPro---ProGlyProGlnGlyLeuProGlyLeuAlaGlyThrAla 859  
QY 1784 GGCAGCCAGTGGGGCAATGGCCCAAGGGGGTGGCGAGCGCTGCCACACATGCTAGTCCTG 1725  
Db 860 GlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeuPro----- 874  
QY 1724 GGGCCAGGATGGGAAGTGGGGGGCCACCTAAGAGGGAAGGGAGGAGATAGACCAGA 1665  
Db 875 -----GlyArg---AspGlySerProGlyGlyLysGlyAspArg----- 886  
QY 1664 CAGGGGACTGGTGGGAGCCAAAGCCCTCCCTCACCCCATCATTTACCAGAACCCCTCC 1605  
Db 887 ---GlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisProGlyProGly 905  
QY 1604 ACGGTTAGGACTGTCTTTCCCAACTCTCACTCTCTCCCAACTCCAGATTCTCTACGCT 1545  
Db 906 -----ProValGlyProAlaGlyLysSer 913  
QY 1544 GGATCTCATAGTCCAGCCAGCTGCGGAGTGTGTCTCCAGGTAGCGGTGAGGT 1485  
Db 914 GlyAspArgGlyGluSerGlyProAlaGlyProAlaGlyAlaProGlyPro----- 930  
QY 1484 CATAGGTTTCTGGATGGAGGG---CCAGGCCCTGGTCCCTGCTGCGATTGAGAGCT 1429  
Db 931 -----AlaGlySerA---GlyAlaProGlyProGlnGlyProArgGlyAspLysGly 947  
QY 1428 GGCAGTCCAGGGAGTCCAGAGCACCGTGCACAGGACGCTAATACATCCCC----- 1378  
Db 948 GluThrGlyGluArgGlyAlaAlaGlyLysGlyLysGlyHisArgGlyPheProGlyAsnPro 967  
QY 1377 -----CACGAGTCCCTGTGGGAGGATGGACGAGAGCATGAGCGGGCCCG 1330  
Db 968 GlyAlaProGlySerProGlyProAlaGlyGlnGlnGlyAlaGlyLysProGlyPro 987  
QY 1329 ---GGCCACACAGGGAAGCAGCTGGCAGCCCTCAGGTGTCTGTCTCCACCTGGGCTG 1273  
Db 988 AlaGlyProArgGlyProVal-----GlyProSerGlyProGlyLysAspGlyThr 1005  
QY 1272 GGTGGGGGTGAGGGT-----GGGACACCATGATGGGAGGAGGAGGATGGAGAGGG 1219  
Db 1006 Ser-GlyHisProGlyProIleGlyProGlyProArgGlyAsnArgGlyGluArgGly 1025  
QY 1218 AAAC-----ACTCTGGCAAGAGGGAGAG 1194  
Db 1025 ySerGluGlySerProGlyHisProGlyGln 1035

## RESULT 13

US-09-579-181-2  
; Sequence 2, Application US/09579181  
; Patent No. 6365372  
; GENERAL INFORMATION:  
; APPLICANT: Yacivuk, John  
; APPLICANT: Yacivuk, Peter  
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)  
; FILE REFERENCE: 16153-4247  
; CURRENT APPLICATION NUMBER: US/09/579,181  
; CURRENT FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: 60/136,620  
; PRIOR FILING DATE: 1999-05-27  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2972  
; TYPE: PRT  
; ORGANISM: Human  
US-09-579-181-2

Alignment Scores:  
Pred. No.: 8,19e-15 Length: 2972  
Score: 318.50 Matches: 318  
Percent Similarity: 31.47% Conservative: 115  
Best local Similarity: 23.11% Mismatches: 476  
Query Match: 3.38% Indels: 467  
DB: 4 Gaps: 61

US-09-931-704-3 (1-5087) x US-09-579-181-2 (1-2972)

QY 63 CCACCTCCGCCAGCTCCCGGAGA----- 86  
Db 803 ProValProLysGlnGlyArgThrValValValAsnAsnProArgAlaProLeu 822  
QY 87 GGAGCGCACCGCGCGCGCGCCAGCCCATGAGCTCCGAGCAGGTGTAACACC 146  
Db 823 GlyProValProValArgProProGlyPro---GluLeuSerAla----- 837  
QY 147 AAACCTAGCCTGTCTTTCATAACATGACAGCAGCGCCCATCTGATACCTAA----- 200  
Db 838 -----GlnProThrProGlyProValProGlnValLeu 848  
QY 201 CCGACCAAGTCACAGCCCTTCAACTCACCTCTGCTCCCGCCAGCTCACACATCTCTTG 260  
Db 849 ProAlaSerLeuMetValSerAlaSerProAlaGlyProProLeuIleProAlaSer--- 867  
QY 261 TGGACTCAACCTCAACCCGCTAAATCAACCAATCCCAAGTCTAACTAATCTGAAC 320  
Db 868 -----ArgProProGlyPro-----ValLeuLeu----- 875  
QY 321 TTTTAAAGTAACCCAGTCTTAAACCTAACTAGCCCAATGCCAATATATATATACCTAG 380  
Db 876 -----ProProLeuGlnProAsnSerGlySerLeuProGlnValLeuProSer 891  
QY 381 CCAAAACCTTAACCTGCTTCCAGTCCAAAGTGTCTCACTGAATCTCAGCTTGTCTCTCA 440  
Db 892 Pro-----LeuGlyValLeuSerGlyThrSerArgProProThrProThrLeuSer 908  
QY 441 CTGAAATCCCAAGAAAGCATATTTCCCACTGCCACATCCCTCTCTACAGCACCAC 500  
Db 909 Leu-----LysProThrProProAlaProValArgLeuSerProAlaProPro 924  
QY 501 CTGCGCTCTGAGCTCTGCTGATCTCTGGATGTCCTCAAACTCTCAGTGCATCAGCCAC 560  
Db 925 ProGly-----SerSerSerLeuLeuLysProLeuThr 935  
QY 561 AAGCCCGACTCGTCAATGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 620  
Db 936 ValProProGlyTyThrPheProProAlaAlaAlaThrThr----- 950  
QY 621 AAAGCCCTCATTTGAAGTCCAACTTTTCCCACTTAACACCAAGACGGGTGAACCTCA 680  
Db 950 ----- 950  
QY 681 CACTGCCACCGTTCCCTGAGAGTGAGACTAAATCTCTCTCAATCTAAACCCACCTTACA 740  
Db 951 -----SerThrThrAlaThrAlaThrThrAlaValProAlaProThr 966  
QY 741 CTTCCCACTCAGGAATCAGATCTCTAGAAATATACCAAACTAAAGCCCATAGGCAGC 800  
Db 967 ---ProAlaProGlnArgLeuLeu-----Ser 975  
QY 801 CCGACCTAGTGGTCTAACCTTATACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 860  
Db 976 ProAspMet-----GlnAlaArgLeuProSerGlyGluValValSerIleGly 991  
QY 861 CGCT 920  
Db 992 GlnLeu-----AlaSerLeuAlaGln-ArgProValAlaAlaAsnAl 1004  
QY 921 TGCTGTCTCCACCCCTCTGACTCCCTCAAGCTCGAGTGGGAGTGGAGAGCTGGCAGGAA 980









Db	60	GlyGluGlnGlyProArgGlyAspArgGlyAspLysGlyGluLysGlyAlaProGlyPro	79
QY	4245	AAAGCATGTGGT-----GGTGGCCCTCAACCCCTGC-CCTCCAGAT	4202
Db	80	ArgGlyArgAspGlyGluProGlyThrLeuGlyAsnProGlyProProGlyPro	99
QY	4201	GTGCCA---CCTGAGGGGCTGGTGGCACTGCCCAAGTGTGTAGGCAAGAGCAGGAGGC	4145
Db	100	ProGlyProProGlyLeuGlyAsnPheAlaAala-----GlnMetAlaGlyGly	117
QY	4144	CATGAGCGCTCGC-----ACCAACCTGAACCACTTCACACTCCCTCGAGCATGACTT	4091
Db	118	PheAspGluLysAlaGlyGlyAlaGlnLeuGlyValMetGlnGlyProMetGlyProMet	137
QY	4091	-----	4091
Db	138	GlyProArgGlyProProGlyProAlaGlyAlaProGlyProGlnGlyPheGlnGlyAsn	157
QY	4090	-----CCTGTAGAAACAGGACAGGACAGGCGCTACTGTACTCTCTCTCC	4049
Db	158	ProGlyGluProGlyGluProGlyValSerGlyPro-MetGlyProArgGlyProProGly	177
QY	4048	CCAGCTCGGTAGACCTTTGGGAGTGGGAG-----GAGACAGGGCTGATCCGATCACA	3995
Db	177	yProGlyLysProGlyAspAspGlyGluAlaGlyLysProGlyLysAlaGlyGluAr	197
QY	3994	CGCCAGCGGTCAGGAAGGCGCAGAGCTCACAGCTTCTGCTCTCGCTGGCTCAACAGG	3935
Db	197	gGlyProProGlyProGlnGlyAlaArgGly-----PheProGlyThrProGly	213
QY	3934	TGTT---GGCATACAGGCTGCTCTCACAAAGTGGAG-CAGGGTTTGAAGGGGAGCG	3879
Db	213	yLeuProGlyValLysGlyHisargGlyTyProGlyLeuAspGlyAlaLysGlyGluAl	233
QY	3878	AAGAGAGAGGTCAAGACTCAGACCCATGAGCCCGCAGGTCCAGGTGACTGCGAGCTG	3819
Db	233	aglyAlaProGlyValLysGlyGluSerGlySer--ProGly-----	246
QY	3818	CTGGAGGCTGATCTTCTTTGAGCGGGTTGAAGTCTTGGCGGAGCGGCCACAGCCAGG	3759
Db	247	---GluAsnGlySerProGlyProMetGlyProArgGlyLeuProGlyGluArgGlyArg	265
QY	3758	TCTGCAGCTCCTTCAGCAGCCAGAACTCGTCCACTCTCTGGAGGAAGTCACTGTGGCAG	3699
Db	266	ThrGlyProAlaGlyAlaAlaGlyAlaArgGlyAsnAspGly-----	279
QY	3698	GGCCAGGAGTCAAGTGGGTTCACTCCAGGACGGCTGGGGAGTGGGTAGCCAGAG	3639
Db	280	-----GlnProGlyProAlaGlyProProGly	288
QY	3638	CTGCCATGACGCCGCAATGCTGCCAGCAGGCGCTCGAGGCTGGTCCAGAAGTGGGCCA	3579
Db	289	-----ProValGlyProAlaGly-----	296
QY	3578	GGCTGGCGGACAGCTCAGCAGTGGCAGCTTGAACGTTGAGGCCACGCAAGTAACAGAA	3519
Db	297	GlyPheProGlyAla-----ProGlyAlaLysGlyGluAlaGlyProThr---	311
QY	3518	GGTGGCTGTAGGCCTCGTATGTTGGGTGACGCGCATTTGTCAATGAGGCTCGCCACA	3459
Db	312	-----GlyAlaArgGly	315
QY	3458	CCTCCAAGTCAACAGTGGCCCTGGGCAGAGTCTCTGCCCCAGCGGGAGGTTGAAGT	3399
Db	316	ProGluGlyAlaGlnGlyProArgGlyGluProGlyThrProGlySerProGlyProAla	335
QY	3398	CTGGCTCGTTGAAGGGGGGCCAGCTAGTGTAGCTGTGTGAAGAGGAGGCTGATGGGA	3339
Db	336	GlyAla-----SerGlyAsnProGlyThrAspGlyLePro-----GlyAlaLysGly	351
QY	3338	AGGAGGAGGCGAGCGCTGGCTCAACCAAGGAGATACCTGCTCCCAAGTTCTTAT	3279
Db	352	SerAtaGlyAlaProGlyAlaGlyAlaProGly--PheProGlyProAr-----	368

Qy	3278	TTTTTGTGTCCTGCACACTGGCCCTGTCTCCATGCTAGGCTTCTTTGTTCATGGCGCTCC	3219
Db	369	-----GlyProPr	371
Qy	3218	CTGAGTCTTCCAAATAAGGATATCATTTGTATGCAGGTGACAGTTGAAGCAGTGGCGGTG	3159
Db	371	o-----	371
Qy	3158	GTGAGCTCTCCAATGGAGAGCGGTAGAAAAAGGAGGAGGAGACGCACAAAGACGCTCA	3099
Db	372	-----AspProGl	374
Qy	3098	GGAAAGCAGTGTGATTTGGCTGGGT-----GGGCAAAACC-----TTGGGAGAGTGCCTC	3051
Db	374	n---GlyAlaThrGlyProLeuGlyProLysGlyGlnThrGlyLysProGlyIleAlaGl	393
Qy	3050	GCTGTGGGGCAGGACAGAGAGAGAGAGAGCTTGAGAGCTGGAGAGCTGGAGAGGTGAAGGTGG	2991
Db	393	yPheLysGlyGluGlnGlyProLysGlyGluProGlyProAlaGlyPro---GlnGlyAl	412
Qy	2990	CCGAGAGAGTCCAGAGACCAAGGA---AGAGGAGTTTCGAAAGGAGGAGCGAGCGCTCAGCA	2934
Db	412	aProGlyProAlaGlyGluGluGlyLysArgGlyAlaArg-----	425
Qy	2933	GTGTGGCTGCAAAAACAGACAGCAAGGAGAGTAATGAGCTTCAAAAACACTGTCTGGGCTTGGCA	2874
Db	426	-----GlyGluPro-----	428
Qy	2873	AGGAGTGCCTCAGCAGCTTGTGAGAGAACATTTGCTGGGGCTGGAGGCAGATTCGAGG	2814
Db	429	-GlyGlyVal-----GlyProIleGlyProProGlyGluArgGlyAlaPr	443
Qy	2813	GGGTAAAGAGTGTGCTGGGTGAGAGGAAGTGGAGCAGAGCGGTGTAGACCACTCCTTGG	2754
Db	443	oGly-----AsnArgGlyPheProGlyGlnAspGly-----LeuAlaGl	456
Qy	2753	AGAAATTTGGCAGGGAACCCAGGAGAGAAAAAGGGTCGGAACTAGAGAGGCGAGTGGGA	2694
Db	456	yPro-----LysGlyAlaProGly-----	462
Qy	2693	GTGAAGCAAAAGGTTTTTCTGCACAAGGAACAAGAGAGAGATGAGAGTGGCCATTTGGAACC	2634
Db	463	-----GluArgGlyProSerGlyLeuAlaGlyPr	472
Qy	2633	A-----GAGCCCATTTGGCGGTGACACACCAGGC---CCACCAAGTGA	2592
Db	472	oLysGlyAlaAsnGlyAspProGlyArgPro---GlyGluProGlyLeuProGlyAlaAr	491
Qy	2591	GGGGCTG---GGGAGAGACGGGTACTCACTAACACTCCGAAAGAGGGTTTCAGCTGATAG	2535
Db	491	gGlyLeuThrGlyArgProGly-----	498
Qy	2534	GAGCAGGGTTGCAGAGGAAGAAATCGAGGCAGTCTCATGAGTAGGACAGGAGGAAGGCT	2475
Db	499	-AspAlaGlyProGlnGlyLysValGlyProSerGlyAlaProGlyGluAspGlyArg-p	518
Qy	2474	ATGGAAGACAGCATGTTATAGGAAGAGGCGAGGCC---TTTCCAAACC	2427
Db	518	roGlyProProGlyProGlnGlyAlaArgGlyGlnProGlyValMetGlyPheProGlyP	538
Qy	2426	TCGGGGAGCGGCTGGGCCCAGGGAAGGGATGAGGG-----	2389
Db	538	roLysGlyAlaAsnGlyGluProGlyLysAla---GlyGluLysGlyLeuProGlyAlaP	557
Qy	2388	-----AGATCTATAACTGGGAAGAGGAGAGCAAGGGGAGAGGC-----	2347
Db	557	roGlyLeuArgGlyLeuProGlyLysAspGlyGlnThrGlyAlaGluGlyProProGlyP	577
Qy	2346	--CTGGGATAGGCAAGGGCCCGAGGGAGGTTCCACAGCTGCAAGCAGCAGAGACAAACT	2289
Db	577	roAlaGlyProAlaGlyGluArgGlyGlu-----GlnGlyAlaP	590



QY 2288 .AATTCTCTGGTGTATCTCTGGGTGGAGCCAGCTCCCGCCACCTTCTTCTCCCTTTTCT 2229  
Db |||||  
590 roGlyProSerGlyPheGlnGlyLeu---ProGlyProGlyProGly----- 606  
QY 2228 TTCCCTGGCTCCAGAGTAGGTCTGGGATGTGGAAGAGGAACACGGGAAGAGGGGGCGG- 2170  
Db |||||  
607 -----GlUGlyGlyLysProGlyAspG 614  
QY 2169 -----GTTCCCATGTCCTCGGCTCTCCCAACCTCTTCTTCTCTCTCTGGGCTCTC 2115  
Db |||||  
614 lnglyValProGlyGluAlaGlyAla-----ProGlyLeuValGlyProA 629  
QY 2114 TGAACACCTCTGGACACCCCAATCTTACAGATTAACCTCCATCCACCCCTAAACC 2055  
Db |||||  
629 rgGlyGluArgGlyPheProGlyGluArgGlySer----- 640  
QY 2054 CCCTGCCACTACTGCTCTGGGTTCCTACAGTCCCTTCCCATGACCCCTAGGAGGC 1995  
Db |||||  
641 -----ProGlyAlaGlnGlyLeuGlnGlyProArgGlyLeuProGlyThrP 656  
QY 1994 CCAGCTAGTAGGG---AGAGGCACACGG-----AAGGAGGCGCCCTGGGTCCAGAGCA 1944  
Db |||||  
656 roGlyThrAspGlyProLysGlyAlaSerGlyProAlaGlyProProGlyAlaGlnGly- 675  
QY 1943 GCAGGGCCCGAGTCTG----- 1927  
Db |||||  
676 -----ProProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyIleAlaGlyP 694  
QY 1926 -----GGGCGAGAGCGCAGC-----TGGAACG 1905  
Db |||||  
694 roLysGlyAspArgGlyAspValGlyGlyLysGlyProGlyGlyAlaProGlyLysAspG 714  
QY 1904 CAGGCATGGAGGTGGGCGCCCTCTGTCCCAACCCCGCCCGGCGCAAGACACCA 1845  
Db |||||  
714 lyGlyArgGlyLeuThrGlyProIleGlyProGlyProAlaGlyAlaAsn----- 731  
QY 1844 CAACACGGTGTAACAGAGGAGACTTCCCGGCCCGCCCGGAGGCGGGGCTGGGCT 1785  
Db |||||  
732 -----GlyGlyLysGlyGluValGly- 738  
QY 1784 GGCAGCCAGTGGGCAATGGGCCAAGCGGGTGGAGCGCTGCCACATGCTAGTCTG 1725  
Db |||||  
739 -----ProProGlyProAlaGlySerAlaGlyAlaArgGlyAla-----ProG 753  
QY 1724 GGGCCAGGATGGGAAGTGGG-----GGGCCACCTAAGA 1689  
Db |||||  
753 lyGluArg---GlyGluThrGlyProProGlyThrSerGlyIleAlaGlyProProGlyA 772  
QY 1688 GGGAGGGCAG-----GAGATAGACCACACAGGGGGACTGG 1653  
Db |||||  
772 laAspGlyGlnProGlyAlaLysGlyGluGlnGlyGluAlaGlyGln-LysGlyAspAla 791  
QY 1652 TGGAGGCCAAGAGCCCTCTCACCCTCATTTACCAAGCCCTCCACGGTTAGGACT 1593  
Db |||||  
792 GlyAlaPro---GlyProGlnGlyProSerGlyAlaProGlyProGlnGly----- 807  
QY 1592 GTCTTTCCCACTCTCTACTCTCTCCCACTCCAGATTCCTACGCTGGATACATCAT 1533  
Db |||||  
808 -----ProThrGlyValThrGlyProLys-----GlyAlaArgGly 819  
QY 1532 AGGTCCCAAGCTGCGGAGTTGGTGTCTCCAGGTAGCGGTGAGGTGCATAGGT----- 1478  
Db |||||  
820 AlaGlnGlyProProGlyAlaThrGlyPheProGly-AlaAlaGlyArgValGlyProPr 839  
QY 1477 -TTTCTGGATGGAGGGCCCGGCTCTGGGTCCCT----- 1444  
Db |||||  
839 oGlySerAsnGly-AsnProGlyProProGlyProProGlyProSerGlyLysAspGlyP 859  
QY 1443 -----GTGCGATTGAGCTGGGAGGTGCCAGAGCGTCACAGGC 1392  
Db |||||  
859 roLysGlyAlaArgGlyAspSerGlyProProGlyArg----- 871  
QY 1391 ACGCTAACATCCCCAC---GAGTCCCTGTGGCAGGATGGACGAGACATGAGGCG 1335

Db 872 --AlaGlyGluProGlyLeuGlnGlyProAlaGlyProProGlyGluLysGlyGluProG 891  
QY 1334 GC---CCGGGCCCCACACGGG---GAAGCAGCTGGCAGCCCTCAGGTGTCTGTCTCCACC 1281  
Db |||||  
891 lyAspAspGlyProSerGlyAlaGluGlyProProGlyProGlnGlyLeuAlaGlyGlnA 911  
QY 1280 TTGGGCTGGGTGGGGGTGAGGTGGACACACAGATGGGAGGAGGAGGATGGAGGAGG 1221  
Db |||||  
911 rgGlyIle-ValGlyLeuProGly-----GlnArgGly---GluArg 923  
QY 1220 GGA-----AACACTCTCTGGCAAGACGGGAGAG 1194  
Db |||||  
924 GlyPheProGlyLeuProGlyProSerGlyGlu 934

Search completed: January 27, 2003, 16:27:31  
Job time : 159.606 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:49:37 ; Search time 24.1925 Seconds  
(without alignments)  
6508.980 Million cell updates/sec

Title: US-09-931-704-4  
Perfect score: 1498  
Sequence: 1 tattattaaagcttcgcgg.....agcccaagtcagctgtgctt 819

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2 1/USPTO spool/US09931704/runat 27012003 154126 3648/app query.fasta\_1.7189  
-DB=PIR\_73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09931704 @CGN\_1\_1 351 @runat 27012003 154126 3648 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 73.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	145	9.7	325	2 T32248	hypothetical prote
2	142.5	9.7	660	1 Q0B3	BHLF1 protein - hu
3	141	9.4	383	2 S22975	gene BCRF2 protein
4	137	9.1	319	2 T32250	hypothetical prote
5	137	9.1	574	2 T43556	Wiskott-Aldrich sy
6	137	9.1	574	2 T38819	wiskott-aldrich sy
7	136.5	9.1	1146	2 A38587	collagen, cornea s
8	135	9.2	371	2 E88633	protein F56B3.1 [i
9	135	9.0	1691	1 S22917	collagen alpha 5(i
10	135	9.0	1763	2 S16366	collagen alpha 2(i
11	134	8.9	302	2 T15936	hypothetical prote
12	134	9.2	627	2 A44112	spidroin 2, dragli
13	133	8.9	304	2 T22482	hypothetical prote
14	132	8.8	1433	2 A46053	bullous pemphigoid

15	132	8.8	1669	1 CGMS4B	collagen alpha 1(I
16	131.5	8.8	303	2 T28999	hypothetical prote
17	131	8.7	317	2 T19143	hypothetical prote
18	131	8.7	754	2 A5267	collagen alpha 5(I
19	130.5	8.7	517	2 T10927	3C3.18c protein -
20	130.5	8.7	677	2 S23296	collagen alpha 2(I
21	130	8.7	660	1 Q0B3	BHLF1 protein - hu
22	130	8.7	1532	2 A61262	collagen alpha 1(X
23	129.5	8.6	304	2 T22602	hypothetical prote
24	129.5	8.6	380	2 T28888	cuticle collagen d
25	129.5	8.6	1670	1 CGHU3B	collagen alpha 3(I
26	129	8.6	316	2 T12888	hypothetical prote
27	129	8.6	576	2 T36729	probable serine/th
28	129	8.6	1147	2 T42627	ADP-ribosylation f
29	128.5	8.6	304	2 T26185	hypothetical prote
30	128.5	8.6	360	2 T26184	hypothetical prote
31	128.5	8.6	360	2 T37285	collagen dpy-2 - C
32	128.5	8.6	688	2 A53330	collagen alpha 2(I
33	128.5	8.6	1541	2 T02831	AAA protein L4171
34	128	8.5	299	2 T19564	hypothetical prote
35	128	8.5	316	2 T19291	hypothetical prote
36	128	8.5	551	2 S7447	HPSII-7 protein -
37	127.5	8.7	316	2 S08169	collagen col-12 pr
38	127.5	8.7	316	2 S08170	collagen col-13 pr
39	127.5	8.5	1250	1 B45219	N-methyl-D-asparta
40	127	8.7	255	2 J00320	hypothetical 24.7K
41	127	8.7	291	2 T26576	hypothetical prote
42	127	8.5	1255	2 T31065	diaphanous protein
43	126.5	8.4	299	2 T29956	hypothetical prote
44	126	8.4	615	2 A05269	collagen alpha 1(I
45	126	8.4	1008	2 T04462	hypothetical prote

## ALIGNMENTS

### RESULT 1

T32248  
hypothetical protein T15B7.4 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000

C/Accession: T32248

R/Pauley, A.; Gattung, S.

submitted to the EMBL Data Library, September 1997

A/Description: The sequence of C. elegans cosmid T15B7.

A/Reference number: Z21139

A/Accession: T32248

A/Status: preliminary; translated from GB/EMBL/DBDJ

A/Molecule type: DNA

A/Residues: 1-325 <PAU>

A/Cross-references: EMBL:AF022985; PIDN:AAB69961.1; GSPDB:GN00023; CBSP:T15B7.4

A/Experimental source: strain Bristol N2; clone T15B7

C/Genetics:

A/Gene: CBSP:T15B7.4

A/Map position: 5

A/Introns: 266/1

C/Superfamily: unassigned collagens

Alignment Scores:

Pred. No.:	0.000832	Length:	325
Score:	145.00	Matches:	73
Percent Similarity:	32.94%	Conservative:	11
Best Local Similarity:	28.63%	Mismatches:	94
Query Match:	9.68%	Indels:	77
DB:	2	Gaps:	11

US-09-931-704-4 (1-819) x T32248 (1-325)

QY	4	TATTAAAGCTTCGCGGAGCGCGCTCGCCCTCCACTCCGCGAGCTCTGGG-----	57
Db	49	TyrGluGlnIleAlaTrpGlnAlaMetIleProThrArgProSerSerGlySerSer	68
QY	58	-----AGAGGA 63	

```

Db 69 PheLeuLeuGlyArgAsnLysArgGlnAlaGlnCysAsnCysGlyAlaGlnSerArgGly 88
Qy 64 CCGGGCCCGCCGGCCCGCC-----CAGCCCATGGGACCTCCGAGCAGGGGACTC 117
Db 89 CysProAlaGlyProProGlyProGlyGlnProGlyAlaProGlyGluGlnGlyHis 108
Qy 118 GTGGGGGATGTAGCTTGCTATGCACGGTGTGTGGCCACCTCC-----162
Db 109 ProGlyLeuAlaGlyGlnProGlySerGlyAlaArgGileAsnProAlaThrGlyArgPro 128
Qy 163 -----TGCACTGCCAGCTCTTAATCGCACAGAGATCCAGGCCCTGGCCC 207
Db 129 GlyPheCysIleThrCysProAlaGlyAla-----ProGlyPro 141
Qy 208 CTCATCCAGAAACCTATGACCTCACCCGCTACCTGAGGAGCATCACTCCGAGCTTAGC 267
Db 142 AlaGlyPro-----ProGlyAlaProGlyProLysGlyAsn 153
Qy 268 TGGGACCTACCTGAACTACCTGCGGCCCTCTTTCACGAGCCTGACTTCAATCCTCCTCG 327
Db 154 -----AsnGlyGlnProGlyAlaProAlaGln-----162
Qy 328 ACTGGGGCAG-----AACTCTGCCAGGCCACGGT 360
Db 163 SerGlyGlyArgGlyProProGlyProArgGlyProAlaGlyAspAlaGlySerProGly 182
Qy 361 CNACTTGGAGTGTGGGAGGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
Db 183 GlnProGlyHisProGlySerProGlyAsnProGlyArgGlyGlyGlnArgSerArgGly 202
Qy 421 GPACAGTCACTCTCTGTGTTACTTGGCTGGCTCAACCTGTCAGGC-----465
Db 203 LeuProGlyProSerGlyArgProGlyProProGlyProAlaGlyGlyProGlyGlnPro 222
Qy 466 TGCCACAGTGAACCTCCGAGTAGCTGCGCCACTCTGTACAGCCTCCAGGCCCTGCT 525
Db 223 GlyHisSerGlyGlyAlaGlySerProGlyPro-----GlnGlyProProGlyProSer 240
Qy 526 GGGCAG-----CATTCAGGTGTCATGCGGACGCTTGGCTACCACTGCGCCAGCCTCT 579
Db 241 GlyGlnProGlyHisSerGlyAsnAspGlyValProGlyAlaProGlyAsnProGlySer 260
Qy 580 GCGAGGAGTGAACCCAGCTGGGCCCTCGCCCTGCCACAGT 622
Db 261 -ProGlyGlyAspAlaLafyrCysProCysProAlaArgSer 274

RESULT 2
QBEB3
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C:Accession: A03742
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A:Reference number: A93065; MUID:85035713; PMID:6092825
A:Accession: A03742
A:Molecule type: DNA
A:Residues: 1-660 <BAN>
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A:Reference number: A03794; MUID:84270667; PMID:6087149
A:Contents: annotation; protein coding region
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
C:Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:
Pred. No.: 0.00131 Length: 660
Score: 142.50 Matches: 92
Percent Similarity: 32.43% Conservative: 16
Best Local Similarity: 27.63% Mismatches: 125
Query Match: 9.75% Indels: 100

```

```

DB: 1 Gaps: 15
US-09-931-704-4 (1-819) x QBEB3 (1-660)
Qy 806 TGGCCCTGGAGGTGGGGTTAAGGCTCAGAGGTGACAGACCATGTGCTCCCAAGTGA 747
Db 159 TrpArgArgSerGlyAla-----GlnArgGlyHisProProGlyAla 174
Qy 746 GGGTGAAGCTGCTGGAGGCTGTCATCTTCTTAAGCCGGTTGAAGTCCTTGGCTG 687
Db 175 GlyGlnArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThr 194
Qy 686 AACGCCATA-----GCCAGTCTGCAGCTCTTTCAGCAGCCAGAGTCAATCCA 639
Db 195 ProAlaAlaProGlyProGlyGlyAlaAlaValProSerGlyAlaThrProHisPro 214
Qy 638 TCCTCTGGAGGAAGTCACTGTGG-----CAGGCCAG 606
Db 215 GluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGln 234
Qy 605 GGGCCAGGCTGGCTCAGTCCCTGGCAGAGGCTGGGCGAGTGGGTAGCCAGCGCTGCCA 546
Db 235 GluProArgLeuProGlnAspLeuAlaAla-----245
Qy 545 TGACACCTGCAATGCTGCCAGCAGCAGC-----519
Db 246 -----GlnArgCysProAlaGlyProProProThrArgSerGlyAlaAlaGln 262
Qy 518 -----CCTGGAGGCTGTACAGAAAGTGGCCAGGCTACGTCGGAGTTGAG 474
Db 263 ArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArg 282
Qy 473 CTGTGGCAG-----CCTGACGTTGAGCCAC-----447
Db 283 ThrTrpArgArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaGlyGlnArg 302
Qy 446 ---GCAAGTAACACAGAGGT-----CACTGTACG 420
Db 303 ProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAla 322
Qy 419 CCTCATAGTTCTGGCTCAGCCGAGCCTGTCATTGAGCGCTTCCGACACACTTCCAAAGTTGA 360
Db 323 ProGlyProGlyGlyAlaAlaVal-ProSerGlyAlaThrProHis-----ProG 340
Qy 359 CCGTGGCCCTGGCAGAGATTTCTGCCCCAGCTCGAGGAGGATT-----GA 315
Db 340 uArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnG 360
Qy 314 AGTCAGGCTCTGTGA-----AGGGGCCCCAGGTAGTTTCAAGTCCAGCTAAGC 261
Db 360 uProArgLeuProGlnAspLeuAlaAlaGlnArgCysProAlaGly-ProProPro 380
Qy 260 TGGCGAGTTGATGCTCCAGGTAGCGGTGAGGTTCATAGTTTCTCGATGGAGGGCCAG 201
Db 380 hrArgSerGlyAlaAlaAlaGlnArgThrHis-----ArgArgProProG 395
Qy 200 GG-----CCTGATCTCT-----GTGGATTAAAGCTGGCA 168
Db 395 lyCysProArgSerAlaArgAsnProGlyCysProArgThrTrpArgArgSerGlyA 415
Qy 167 CTGAGGAGGTGTCACAGCAGCCTGTCATAGTCAAGCTAACATCCCCAGGAGTCCCTG 108
Db 415 la-GlnArgGlyHisProProProGlyAlaGlyGlnArgProSer-----429
Qy 107 CTCGAGGTTCATGGGGCTGGGGCGCGGCCCGCGCGCTCTCTCCAGAGGCT 48
Db 430 -----GlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAla 447
Qy 47 GCGGAGGTGGAGGCGGAGCCCGGCTCGCGG 15
Db 448 ProGlyProGlyGlyAlaAlaValProSer 458

RESULT 3

```

\_\_\_\_\_





```
QY 409 GAACTATGAGCGGTACAGTCACCTCTCTGTGTTACTTGTGCTGCTCAACCGTCAGGCTGC 468
Db 362 GlyLeuThrGlyProGlnGlyProProGlyLeuProGlyAsnProGlyArgProGlyAla 381
QY 469 CAC-----AGCTGAACCTCCGACG 486
Db 382 LysGlyGluProGlyAlaProGlyLysValIleSerAlaGluGlySerSerThrIleAla 401
QY 487 TAGCTGCGCCCACTTCTGTACAGCTCCAGGCGCTGCTGGCAGCATTTGCAGGTCTCAT 546
Db 402 LeuProGlyPro-----ProGlyProProGlyProIleGly----- 413
QY 547 GCGCAGCTTGCTACCTCCAGCTCCAGCTCCAGGCGCTGCTGGCAGCATTTGCAGGTCTCAT 603
Db 414 -----ProThrGlyPro-ProGlyValProGly-----ProVal-GlyP 426
QY 604 CCTGCGCCCTGCCACAGTCACTTCTCCAGAAGATGATGACTTCTGCTGCTGAAGGA 663
Db 426 roAlaGlyLeuPro----- 430
QY 664 GCTGACAGCTGGCTATGCGGTTTCAG---CCAAGGACTTCAACCGGCTTAAGAGAAGAT 720
Db 431 -----GlyGlnGlnGlyProArgGlyGluLysGlySerAlaValGluV 445
QY 721 GCAGCTCCAGCAGCTTCACTCACCTCCAGCT-----TGGAGGCACATGTTTCTTGACC 774
Db 445 alValIleGluThrIleValSerSerLeuAlaSerGlnMetLeuSerAspL 465
QY 775 TCTGACCTTAACTCCACACCTCCAGGCCCA 806
Db 465 euGlnGlyArgAlaGlyProProGlyProProGlyPro 477
RESULT 8
E88633
protein F56B3.1 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: E88633
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E88633
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-371 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAO2612.1; PID:g2854198; GSPDB:GN00022; CESP:F56B3.1
A;Note: contains similarity to collagens
C;Genetics:
A;Gene: F56B3.1
A;Map position: 4
C;Superfamily: unassigned collagens
Alignment Scores:
Pred. No.: 0.00531 Length: 371
Score: 135.00 Matches: 71
Percent Similarity: 35.68% Conservative: 10
Best Local Similarity: 31.28% Mismatches: 99
Query Match: 9.23% Indels: 48
DB: 2 Gaps: 12
US-09-931-704-4 (1-819) x E88633 (1-371)
```

```
QY 518 CTTGGAGGCTGTATCAGAGCTGGCCAGGCTACCTCGAGTTTCAG----- 474
Db 144 Pro-----ArgGlyProProGlyGlnAlaGlyLeuAspGlyLeuProGlyAla 159
QY 473 CTTGGCAGCTGACGGTTGAGGCCAGCAAGTAACACAGAGGCTGACTGTACGCTCAT 414
Db 160 ProGlyGlnProGlySerAsnGlyGlyAlaGlySerAsnGlyAla----- 174
QY 413 AGTTCTGGGTCCAGCCGACGCTGTCATTGAGGCTTCGCCACACTTCCAAAGTTGACCGTGG 354
Db 175 SerGluGlySerAlaGlyGlyCys-----LysThrCysProAlaGlyPro--- 189
QY 353 CCCTGGCGCAGAGTTTCTGCCCCCAGTCCAGGAGGATTGAAGTCAGGCTCGTTGAAGGGG 294
Db 190 -----ProGlyProProGlyProAlaGlyGlnAlaGlyArgProGly 203
QY 293 GCCCCAGGTAGTTCAGTAGTCCAGCTAAGCTGCGGAGTTGATGCTCCAGGTAGCGGG 234
Db 204 AsnAspGlyGlnProGlyAlaProSerPheGlyGlyValGlyAlaProGlyAlaPro 223
QY 233 TGAGGTTCATAGTTTCTGATGGAGGGGCCAGG-----CCTGGATCTCTCTGCGCA 181
Db 224 GlyProAlaGlyAspAlaGlySerProGlyGlnProGlyAlaProGlyGlnProGlyArg 243
QY 180 TTAAGAGCTGGCACTGCGAGGAGGTGCCACAGCACCGTGCATAGCGCAAGCTAACATCCCC 121
Db 244 ProGlyLysAsnAla-GlnGlyGlySerSerArgProGlyProProGlyProAlaGlyPr 263
QY 120 CACAGAGTCCC-----CTGCTCGAGGTCCTATGGGGCTGGGGCCG----- 81
Db 263 oProGlyProProGlyAsnAsnGlyAlaProGlyGlyGlyTyrGlyValGlyProProGly 283
QY 80 -----CGCCGCGCGGGCGGCTCTCTCCAGAGGCTGGCGGAGTGGG 37
Db 283 yProProGlyProSerGlyArgProGlyAla-----ProGlyGlnProGlyProAs 300
QY 36 AGGCGAGCGCGCGGCTCCG 18
Db 300 pGlyGlnProGlyAlaPro 306
RESULT 9
S22917
collagen alpha 5(IV) chain precursor, renal splice form - human
N;Alternate names: procollagen alpha 5(IV) chain
N;Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 27-Feb-1997 #text_change 21-Jul-2000
C;Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598;
R;Zhou, J.; Hertz, J. M.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 267, 12475-12481, 1992
A;Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and ide
n Alport syndrome patient.
A;Reference number: S22917; MUID:92316923; PMID:1352287
A;Accession: S22917
A;Molecule type: mRNA
A;Residues: 1-967 <ZHO>
A;Cross-references: GB:M90464; NID:g180826; PIDN:AAA52046.1; PID:g553234
R;Zhou, J.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 269, 6608-6614, 1994
A;Title: Structure of the human type IV collagen COL4A5 gene.
A;Reference number: A54365; MUID:94165049; PMID:8120014
A;Accession: A54365
A;Molecule type: DNA
A;Residues: 1-922 <ZH2>
A;Cross-references: GB:U04470; NID:g463378; GB:U04520; NID:g463428; PIDN:AAO27816.1;
R;Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paep, A.; Tryggvason, K.
Science 261, 1167-1169, 1993
A;Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited
A;Reference number: A57079; MUID:93361972; PMID:8356449
A;Accession: A57079
A;Molecule type: DNA
A;Residues: 1-27 <ZH4>
```





```

Db 1339 -----GlyMetLysGlyProSerGlyValPro 1347
QY 385 CAATGACAGCGTCGGCTGACCCAGAACTATGAGCGTACAGTCACCTCTGTGTACTT 444
Db 1348 -----GlySerAlaGlyProGlu-----GlyGluProGlyLeuLeu 1359
QY 445 GCGTGCCTCAACCGTCAGGCTGCCACAGCTGAATCCGACGTAGCCTGGCCCACTTCTG 504
Db 1360 GlyProGlyProGlyLeu----- 1367
QY 505 TACCAGCTCCAGGCTGCTGGCGAGCATTCAGAGTGTATGCGCAGCTTGGCTACCC 564
Db 1368 -----ProGlyProSerGlyGlnSerIleIleIleLysGlyAspAla-GlyProPr 1384
QY 565 A-----CTGCCCCAGCCTCT 579
Db 1384 OGlylleProGlyGlnProGlyLeuLysGlyLeuProGlyProGlnGlyProGlnGlyLe 1404
QY 580 GCCAGGACTGAGCCAGCCTGGCGCCCTGGCCCTGCCACAGTGAATTCCTCCAGAAAGAT 639
Db 1404 uProGly-----ProThrGlyProProGlyAspProGlyArgAsnGlyLeuProGlyPh 1422
QY 640 GGATGACTTCTGCTGCTGAGAGCTGACAGCTGGCTATGGCTTCAGCCAGGACTT 699
Db 1422 eAspGly-AlaGlyGly---ArgLysGlyAspProGlyLeuProGlyGlnProGlyThr- 1440
QY 700 CAACCGCTTAAGAAGAAGATGCAGCTCCAGCAGCTTCAGTCACCTGCATTCGGAGGC 759
Db 1441 --ArgGlyLeuAspGlyProProGlyProAspGlyLeuGlnGlyPro----- 1455
QY 760 ACATGTTTCTGACCTCTGACCTTAAACCCACACCTCCAGGCCCACTGAGCTGTGCT 818
Db 1456 -----ProGlyProProGlyThrSerSerValAla 1465

RESULT 10
S16366
collagen alpha 2(IV) chain precursor - pig roundworm
C:Species: Ascaris suum (pig roundworm)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: S16366
R:Pettitt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A:Title: The complete primary structure of a nematode alpha-2(IV) collagen and the parti
A:Reference number: S16366; MUID:91340768; PMID:171907
A:Accession: S16366
A:Molecule type: mRNA
A:Residues: 1-1763 <JB1>
A:Cross-references: GB:M67507; NID:gl59648; PIDN:AAA18014.1; PID:gl59649
C:Genetics:
A:Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfid
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F:27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F:43-1525/Domain: collagenous #status predicted <COL>
F:197-199/Region: cell attachment (R-G-D) motif
F:1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1530-1638/Domain: repeat NC1 #status predicted <NC12>
F:1639-1763/Domain: repeat NC1 #status predicted <NC12>
F:31.34.39.41.536.539/Disulfide bonds: interchain #status predicted
F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1593-1599,1702-1709/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.: 0.00515 Length: 1763
Score: 135.00 Matches: 84
Percent Similarity: 34.13% Conservative: 16
Best Local Similarity: 28.67% Mismatches: 102
Query Match: 9.01% Indels: 91
DB: 2 Gaps: 15

```

```

US-09-931-704-4 (1-819) x S16366 (1-1763)
QY 19 GGAGCGCGGCTCGCCCTCCCACTCCGACGCTCTGGAGAGAGAGCCGCGCCGCGCGG 78
Db 870 GlyLeuAlaGlyAlaProGlyPheProGlyAlaLysGlyGluProGlyLeu 889
QY 79 CCCGCCCCAG---CCCCATGGACCTCCGAGCAGG-----GGA 114
Db 890 ProGlyLysGlyGluProGlnGlyProGlyGlnProGlyAlaProGlyPheProGly 909
QY 115 CTCGTGGGGATGTAGCTTGCCT-----ATGCACCGTGTCTGTG 153
Db 910 GlnLysGlyAspGluGlyLeuProGlyLeuProGlyValSerGlyMetLysGlyAspThr 929
QY 154 GCACCTCCCTCAGTGCAGCTCTTAATCGCACAGAGATCCAGGC---CCTGGCCCTC 210
Db 930 Gly-LeuProGlyValProGlyLeu-----AlaGlyProProGlyGlnProGly-PheP 947
QY 211 CATCCAGAAAACCTATGACCTCACCCGCTACTGTGAGCATCAACTCCGAGCTTAGCTGG 270
Db 947 roGlyGlnLysGlyGln-Pro---GlyPheProGlyValAlaGlyAlaLysGlyGluAla 965
QY 271 GACCTACTGAACTACCTGGGGCCCTTTCAACGAGCTGACTTCAATCCTCCTCGACT 330
Db 966 GlyLeuProGlyLeuProGlyAlaPro----- 974
QY 331 GGGGGCAGAAACTCTGCCCGAGGCCACCGTCAACTTGAAGTGTGGCGAAGCCTCAATGA 390
Db 975 ---GlyGlnLysGlyGluGlnGly----- 981
QY 391 CAGCTCGGGTGAGCCCAAGAACTATAGCGGTACAGTCACTCTGTGTACTTGGTGG 450
Db 982 ---LeuAlaGlyLeuProGlyIleProGlyMetLysGlyAlaProGlyIleProGlyAla 1000
QY 451 CCTCAACCGTCAGCTGCCACAGCTGAATCCGAGTACCTCGAGTACCTGGC----- 495
Db 1001 ProGlyGlnAspGly-----LeuProGlyLeuProGlyValLysGlyAspArg 1016
QY 496 -----CCACTTCTGTATCCAGCGCTCCAGCGCTCTGGGAGCATTCAGG 540
Db 1017 GlyPheAsnGlyLeuProGlyGluLysGlyGluProGlyProAlaAlaArgAspGlyGlu 1036
QY 541 TGTATGCGCAGCTGTGGTACCACCTCCAGCTCCGAGCTCTGCCAGGAGCTAG---CCAGC 597
Db 1037 ---LysGlyGluProGlyLeuProGlyGlnProGly-LeuArgGlyProGlnGlyProPr 1055
QY 598 CTGGGCCCTGGCCCTGCCACAGTACTTCTCCAGAGATGGATGATCTTCTGGCTGCT 657
Db 1055 oGlyLeuProGly-LeuProGlyLeuLysGlyAspGluGly----- 1068
QY 658 GAAGGAGCTGCAGACCTGGCTATGGCTTCAGCCAAAGACTTCAACCGCTTAAGAAGAA 717
Db 1069 -----GlnProGlyTyrgly-----AlaProGlyLeuMetGlyC 1080
QY 718 GATCAGCTCCAGCAGCTTCAGTCACTCCCTGCATCCCTGAGGAGGACATGTTTCTGACCTCT 777
Db 1080 LuLysGlyLeuProGlyLeuProGlyLys----- 1089
QY 778 GACCTTAACCCACACCTCCAGGCCCA 806
Db 1090 --ProGlyArgProGlyAlaProGlyPro 1098
RESULT 11
T15936
hypothetical protein EGAP7.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T15936
R:Miller, N.
submitted to the EMBL Data Library, May 1996
A:Description: The sequence of C. elegans cosmid EGAP7.
A:Reference number: Z18433
A:Accession: T15936

```



```
Db 373 GlnGlnGlyProSerGlyPro-----GlySerAlaSerAlaAlaAlaAlaAla 390
QY 104 GGAGGTCCTCATGGCGTGGGGCGCGCGCGCGCGCGCGCG-----GCTCTCTCCCA 54
Db 391 AlaGlyProGlyGlyTyrglyProGlyGlnGlnGlyProGlyGlyTyraProGlyGln 410
QY 53 GAGGCT---GGCGGAGTGGGAGGCGGAGCGCGCGCTCCGCGC 15
Db 411 GlnGlyProSerGlyProGlySerAlaSerAlaAlaAlaAla 424
RESULT 13
T22482
hypothetical protein F52B11.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T22482
R;Matthews, L.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19569
A;Accession: T22482
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-304 <WIL>
A;Cross-references: EMBL:282268; PIDN: CAB05195.1; GSPDB: GN00022; CESP: F52B11.4
A;Experimental source: clone F52B11
C;Genetics:
A;Gene: CESP: F52B11.4
A;Map position: 4
A;Introns: 27/3
C;Superfamily: unassigned collagens
Alignment Scores:
Pred. No.: 0.00772 Length: 304
Score: 133.00 Matches: 60
Percent Similarity: 32.27% Conservative: 11
Best Local Similarity: 27.27% Mismatches: 68
Query Match: 8.88% Indels: 81
DB: 2 Gaps: 12
US-09-931-704-4 (1-819) x T22482 (1-304)
QY 16 GCGGAGCGCGGCTCGCCCTCCACTCCGCCAGCTCTGGAGAGAGCGCGCGCGC 75
Db 110 AlaGlyProAlaGlyThrProGlyLysProGlyArgProGlyArgProGlyAlaAlaGly 129
QY 76 CGGCGCGCGCGCGCCATGACCTCCGAGCAGGAGGAGCTCTGGGGGATGTAGCTTG 135
Db 130 LeuProGly---AsnProGlyArgProProAlaGln----- 140
QY 136 CCTATGACAGGTGCTGTGGCACTCCCTGCAGTGCCAGCTTTAATGCGCACAGGATCC 195
Db 141 ProCysGluProIleThrProProCys----- 150
QY 196 AGGCGCTGGCCCTCCATCCAGAAAACCTATGACCTCACCGGTACCTGAGCATCAACT 255
Db 151 LysProCysProGlnGlyProAlaGly-----AlaProGlyAlaProGly 165
QY 256 CCGCAGCTTAGCTGGAGCACTACTGTAACCTGCGGGGCGCCCTTTCAACAGCGCTGACTT 315
Db 166 ProGlnGlyAspAlaGlyAlaProGlyAlaProGly----- 177
QY 316 CAATCTCTCTCGACTGGG-----GSCAGAAACTCTGCCAGC 351
Db 178 GlnGlySerGlyAlaGlyAlaProGlyProAlaGlyProLysGlyAlaSerGlyAlaPro 197
QY 352 GGCAC---GGTCAACTTGAAGTGTGGGAGACCTCAATGACAGGCTCGCGGTGACCCA 408
Db 198 GlyAsnProGlyGlnAlaGlyAlaProGlyGlnPro-----GlyAlaAspAla 213
QY 409 GAACATATGAGCGGTACAGTCACTCTCTGTGTACTTGGCTGGGCTCAACCGTCAGGCTGC 468
Db 214 GlnSerGluSerIleProGlyAlaPro-----GlyGlnAlaGlyProGlnGly--- 229
```

```
QY 469 CACAGCTGAATCCGACGTAGCTGGCCCACTTCTGTACACGACCTCCAGGCGCTGCTGGG 528
Db 230 -----ProGlyProAlaGly 235
QY 529 CAGCATTCACAGGTGTATGGCGACGCTTGGCTACCCACTGCCCCAGCCCTCTCCAGGGAC 588
Db 236 Ser-----ProGlyAlaProGlyGly-ProGlyGly 245
QY 589 TGAG-----CCAGCGCTGGGCGCCCTGGCCCTGCCACACAGTGCAC 625
Db 245 nProGlyAlaProGlyGlnLysGlyProSerGlyAlaProGlyGlnProGlyAlaAsp 264
RESULT 14
A46053
bullous pemphigoid antigen, BPAG2, type XVII collagen alpha 1-chain - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-Sep-1993 #sequence_revision 21-Sep-1993 #text_change 05-Nov-1999
C;Accession: A46053
R;Li, K.; Tamai, K.; Tan, E.M.L.; Uitto, J.
J. Biol. Chem. 268, 8825-8834, 1993
A;Title: Cloning of type XVII collagen. Complementary and genomic DNA sequences of the
segment, and unusual features in the 5'-end of the gene and the 3'-untranslated region
A;Reference number: A46053; MUID: 93232041; PMID: 8473327
A;Accession: A46053
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1433 <LI1>
A;Cross-references: GB: L08407; NID: G309182; PIDN: AAA37443.1; PID: G309183
A;Note: sequence extracted from NCBI backbone (NCBIN: 129627, NCBIP: 129628)
Alignment Scores:
Pred. No.: 0.00902 Length: 1433
Score: 132.00 Matches: 99
Percent Similarity: 27.85% Conservative: 11
Best Local Similarity: 25.06% Mismatches: 103
Query Match: 8.81% Indels: 182
DB: 2 Gaps: 22
US-09-931-704-4 (1-819) x A46053 (1-1433)
```

```
QY 37 CCCACTCGCCAGCTCTGGG-----AGAGAGCGCGC----- 69
Db 642 ProGlyProProGlySerGlyGlyLysGlyAspArgGlyIleAlaGlyGlnGlyPro 661
QY 70 -----CCGCGCGCGCGCGC-----CCCCAG 90
Db 662 GlnGlyLeuProGlyValProGlyProProGlyLeuArgGlyHisSerGlySerProGly 681
QY 91 CCCCATGACCTCCGAGC-----AGGGGACTCGTGGGGGATGTAGCTTGCCT 138
Db 682 ProGlnGlyProProGlyAlaValGlyProGlnGlyLeuArgGlyAspValGlyLeuPro 701
QY 138 ----- 138
Db 702 GlyValLysGlyAspLysGlyLeuMetGlyProProGlyProLysGlyAspGlnGlyGlu 721
QY 139 -----ATGACGCTGCTGTGGCA 156
Db 722 LysGlyProArgGlyLeuThrGlyGluProGlyIleArgGlyLeuProGlyAlaValGly 741
QY 157 CCTCCCTGCAGCT-----GCCAGCTCTTAATCGCACAGGAGATCCAGG 198
Db 742 GluProGlyAlaLysGlyAlaMetGlyProAlaGlyAlaAspGlyGlnGlnGlySerArg 761
QY 199 -----CCCTGGCGCCCTCCATCCAGAAAACCTATGACCTCAC 234
Db 762 GlyGlnGlnGlyLeuThrGlyMetProGlyThrArgGlyPro-----ProGly 777
QY 235 CCGCTA-----CCTGGAGCATCAACTCCGCGAGCTTAGCTGGGACCTACCT 279
Db 778 ProAlaGlyAspProGlyLysProGlyLeuThrGlyProGlnGlyProGlnGlyLeuPro 797
```

QY 280 GAACACTACCTGGGCGCCCTTCAACGAGCCTGACTTCAATCCTCCTCGACTGGGGCAGA 339  
 Db 798 GlySerProGlyArgPro-----GlyThrLysGlyGlu 808  
 QY 340 AACTCTGCCAGGCCAC----- 357  
 Db 809 ProGlyAlaProGlyArgValMetThrSerGluGlySerSerThrIleThrValProGly 828  
 QY 358 -----GCTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGCGCTGCGCT 402  
 Db 829 ProProGlyProProGlyAlaMetGlyProProGlyProProGlyThrProGlyProAla 848  
 QY 403 GACCCAGAA---CTATGAGCGGTACAGTCACT----- 432  
 Db 849 GlyProAlaGlyLeuProGlyGlnGlnGlyProArgGlyGluProGlyLeuAlaGlyAsp 868  
 QY 433 -----CCTGTGTTACTTGGCTGG----- 450  
 Db 869 SerPheLeuSerSerGlySerSerIleSerGluValLeuSerAlaGlnGlyValAspLeu 888  
 QY 451 -----CCTCAACCGCTGAGCTGCCACAGCTGAACCTCCGAGTACGCTGGCCCACTT 501  
 Db 889 ArgGlyProProGlyProProGlyProArgGly-----Pro-----ProGlyProSer 904  
 QY 502 CTGTACACGCTCCAGGCGCTGCTGGGCACATGTCAGGTGTCATGCGCAGCGCTGGCTA 561  
 Db 905 IleProGlyProProGlyProArgGlyPro-----ProGlyGluGly---Val 919  
 QY 562 CCCACTGCCCGCCGCTCTGCCAGG-----ACTGA 591  
 Db 920 ProGlyProProGlyPro---ProGlySerPheLeuThrAspSerGluThrPhePheThrG 939  
 QY 592 GCCAGCTGGGCGCCCTGGCCCTGCCACAGTGACTTCTCCAGAGATGATGACTTCTG 651  
 Db 939 yProProGlyProProGlyPro---ProGlyProLysGlyAspGlnGly----- 954  
 QY 652 GCTGTGAAGGAGCTGCACAGCTGGCTATGGCTTACGCCAAGACTTCAACCGCTTAA 711  
 Db 955 -----AspProGlyValProGlyThrProGlyIleSerGlyGlyLeuS 969  
 QY 712 GAAGAAGATGCAGCTCCAGCAGCTT-----CAGTCACCCCTGCAGCTTGGAGGCACATGG 765  
 Db 969 eHisGlyAlaSerSerSerThrLeuTyMetGlnGlyPro----- 982  
 QY 766 TTTCTGACCTCTGACCTTAAACCCACACCTCCAGGCCCA 806  
 Db 983 -----ProGlyProProGlyProProGlyPro 991  
 RESULT 15  
 COMS4B  
 collagen alpha 1(IV) chain precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 16-Jun-2000  
 C:Accession: A33525; S01454; A28066; A02864; A25636; A29301; S19079; A32003; A31766; S19079  
 R:Muthukumar, G.; Blumberg, B.; Kurkinen, M.  
 J. Biol. Chem. 264, 6310-6317, 1989  
 A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Diff  
 A:Reference number: A33525; MUID:89197932; PMID:2703490  
 A:Accession: A33525  
 A:Molecule type: mRNA  
 A:Residues: 1-1669 <MUT>  
 A:Cross-references: EMBL:J04494; NID:G556296; PIDN:AAA50292.1; PID:G556297  
 R:Wood, L.; Theriault, N.; Vogeli, G.  
 FEBS Lett. 227, 5-8, 1988  
 A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the al  
 A:Reference number: S01454; MUID:88112221; PMID:3338568  
 A:Accession: S01454  
 A:Molecule type: mRNA  
 A:Residues: 1-185; 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-71  
 A:Cross-references: EMBL:X06777  
 R:Killen, P.D.; Burbello, P.; Sakurai, Y.; Yamada, Y.  
 J. Biol. Chem. 263, 8706-8709, 1988  
 A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain

A:Reference number: A28066; MUID:88243724; PMID:3379041  
 A:Accession: A28066  
 A:Molecule type: mRNA  
 A:Residues: 1-129 <K11>  
 A:Cross-references: EMBL:J03758; NID:G192869; PIDN:AAA37439.1; PID:G192670  
 R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Vos  
 Eur. J. Biochem. 147, 217-224, 1985  
 A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alph  
 A:Reference number: A02864; MUID:85127033; PMID:2578961  
 A:Accession: A02864  
 A:Molecule type: mRNA  
 A:Residues: 1276-1669 <OBE>  
 A:Cross-references: EMBL:X02201; NID:G50233; PIDN:CAA26132.1; PID:G1333876  
 R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.  
 Gene 43, 301-304, 1986  
 A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligod  
 A:Reference number: A25636; MUID:86301886; PMID:3755692  
 A:Accession: A25636  
 A:Molecule type: mRNA  
 A:Residues: 1149-1396, 'S', 1398-1424 <NAT>  
 A:Cross-references: EMBL:M14042; NID:G192286; PIDN:AAA37342.1; PID:G192287  
 A:Note: the authors translated the codon CAG for residue 1374 as Arg  
 R:Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pih  
 J. Biol. Chem. 262, 8496-8499, 1987  
 A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(I  
 A:Reference number: A94680; MUID:87250460; PMID:3597383  
 A:Accession: A29301  
 A:Molecule type: mRNA  
 A:Residues: 1441-1669 <KUR>  
 A:Cross-references: EMBL:M15832; NID:G192282; PIDN:AAA37340.1; PID:G387115  
 R:Killen, P.D.; Burbello, P.D.; Martin, G.R.; Yamada, Y.  
 J. Biol. Chem. 263, 12310-12314, 1988  
 A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequ  
 A:Reference number: S19079; MUID:88315019; PMID:2842328  
 A:Accession: S19079  
 A:Molecule type: DNA  
 A:Residues: 1-28 <K12>  
 A:Cross-references: EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PID:G466503  
 R:Kayes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.  
 J. Biol. Chem. 263, 19274-19277, 1988  
 A:Title: Head-to-head arrangement of murine type IV collagen genes.  
 A:Reference number: A92702; MUID:89066738; PMID:3198626  
 A:Accession: A32003  
 A:Molecule type: DNA  
 A:Residues: 1-28 <KAY>  
 A:Cross-references: EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PID:G450449  
 R:Burbello, P.D.; Martin, G.R.; Yamada, Y.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988  
 A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional pr  
 A:Reference number: A94220; MUID:89071759; PMID:3200851  
 A:Accession: A31766  
 A:Molecule type: DNA  
 A:Residues: 1-28 <BUR>  
 A:Cross-references: EMBL:M23333; NID:G340878; PIDN:AAA51625.1; PID:G535668  
 R:Sakurai, Y.; Sullivan, M.; Yamada, Y.  
 J. Biol. Chem. 261, 6654-6657, 1986  
 A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen ge  
 A:Reference number: S19094; MUID:86196099; PMID:3009468  
 A:Accession: S19094  
 A:Molecule type: DNA  
 A:Residues: 1110-1135; 1189-1316; 1342-1383; 1418-1487 <SAK>  
 A:Cross-references: EMBL:M13027  
 R:Schuppan, D.; Timpl, R.; Glangville, R.W.  
 FEBS Lett. 115, 297-300, 1980  
 A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane  
 A:Reference number: S16909; MUID:80246483; PMID:6772473  
 A:Accession: S16909  
 A:Molecule type: protein  
 A:Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-1  
 R:Schuppan, D.; Glangville, R.W.; Timpl, R.  
 Eur. J. Biochem. 123, 505-512, 1982  
 A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial a  
 A:Reference number: A25991; MUID:82186723; PMID:6804236



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:44:12 ; Search time 10.63 Seconds  
(without alignments)  
6391.168 Million cell updates/sec

Title: US-09-931-704-4  
Perfect score: 1498  
Sequence: 1 tattattaagatttcgcgg.....agccacagtcagctgtgctt 819

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cn2 1/USPTO spool/US09931704/runat 27012003 154125 3623/app query.fasta 1.7189  
-DB=SwissProt 40 -OPMT=fastan -SUFFIX=rep -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09931704 @CGN 1 1.141 @runat 27012003 154125 3623 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145.5	9.7	684	CA39 HUMAN	Q14050 homo sapien
2	142.5	9.7	660	YH11 EBV	P03181 epstein-bar
3	137	9.1	1685	CA54 HUMAN	P29400 homo sapien
4	135	9.0	1763	CA24 ASCSU	P27393 ascaris suu
5	134	9.2	627	SPD2 NEPC	P46804 nephila cla
6	133	8.9	1083	T2D3 HUMAN	O0268 homo sapien
7	132	8.8	1669	CA14 MOUSE	P02463 mus musculus
8	131	8.7	317	YQ35 CAEEL	Q09456 caenorhabdi
9	131	8.7	754	CA54 CANFA	Q28247 canis famil
10	130	8.7	660	YH11 EBV	P03181 epstein-bar
11	129.5	8.6	360	CCD2 CAEEL	P35799 caenorhabdi
12	128	8.5	316	CC07 CAEEL	P28630 caenorhabdi
13	127.5	8.7	316	CC12 CAEEL	P20631 caenorhabdi
14	127.5	8.5	1237	CC13 CAEEL	Q00961 rattus norv
15	127.5	8.5	1237	NME3 RAT	Q01955 homo sapien
16	127.5	8.5	1670	CA34 HUMAN	Q08808 mus musculus
17	127	8.5	1255	D1A1 MOUSE	P34687 caenorhabdi
18	126.5	8.4	299	CC34 CAEEL	

19	126	8.4	1069	1	S24B ARATH	Q9M081 arabidopsis
20	125.5	8.4	415	1	SYN1 CANFA	O62732 canis famil
21	125	8.3	721	1	YK82 MYCTU	Q10690 mycobacteri
22	125	8.3	1362	1	CA21 CHICK	P02467 gallus gall
23	124.5	8.3	494	1	ATF7 HUMAN	P17544 homo sapien
24	124.5	8.3	1806	1	CA1B HUMAN	P12107 homo sapien
25	124	8.3	779	1	CA11 BOVIN	P02453 bos taurus
26	124	8.5	825	1	ICP0 HSV2H	P28284 herpes simp
27	124	8.5	1147	1	MYSB ACACA	P19706 acanthameb
28	124	8.3	1804	1	CA1B MOUSE	Q61245 mus musculu
29	124	8.3	2944	1	CA17 HUMAN	Q02388 homo sapien
30	123.5	8.2	680	1	CA1A MOUSE	Q05306 mus musculu
31	123.5	8.2	743	1	CA18 MOUSE	Q00780 mus musculu
32	123.5	8.2	3530	1	MY15 HUMAN	Q9UKN7 homo sapien
33	123	8.2	680	1	CA1A HUMAN	Q03692 homo sapien
34	122.5	8.2	453	1	CA44 BOVIN	Q29442 bos taurus
35	122	8.1	705	1	SYN1 HUMAN	P17600 homo sapien
36	122	8.1	1460	1	CA11 CANFA	Q9X9J7 canis famil
37	122	8.1	1464	1	CA11 HUMAN	P02452 homo sapien
38	121.5	8.1	296	1	PRP3 MOUSE	P05143 mus musculu
39	121.5	8.1	555	1	GPI CHURE	Q9FDP6 chlamydomon
40	121.5	8.1	744	1	CA18 RABIT	P14282 oryctolagus
41	121.5	8.3	1262	1	CA13 CHICK	P12105 gallus gall
42	121.5	8.1	1838	1	CA15 HUMAN	P20908 homo sapien
43	120.5	8.0	475	1	S3A2 MOUSE	Q62203 mus musculu
44	120.5	8.0	674	1	CA1A CHICK	P08125 gallus gall
45	120.5	8.0	1567	1	FMN2 MOUSE	Q9J104 mus musculu

## ALIGNMENTS

## RESULT 1

CA39 HUMAN	STANDARD:	PRT:	684 AA.
ID CA39 HUMAN	Q14050; Q9UPE2; Q9H4G9; Q13681;		
AC Q14050; Q9UPE2; Q9H4G9; Q13681;			
DT 15-JUN-2002 (Rel. 41, Created)			
DT 15-JUN-2002 (Rel. 41, Last sequence update)			
DT 15-JUN-2002 (Rel. 41, Last annotation update)			
DE Collagen alpha 3(IX) chain precursor.			
GN COL9A3.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=Cartilage;			
RX MEDLINE=96163887; PubMed=8586434;			
RA Brewton R.G., Wood B.M., Ren Z.-X., Gong Y., Tiller G.E., Warman M.L.,			
RA Lee B., Horton W.A., Olsen B.R., Baker J.R., Mayne R.;			
RT "Molecular cloning of the alpha 3 chain of human type IX collagen:			
RT linkage of the gene COL9A3 to chromosome 20q13.3.";			
RL Genomics 30:329-336 (1995).			
RN [2]			
RP SEQUENCE FROM N.A., AND VARIANTS 563-GLY--PRO-565 DEL AND			
RP 564-PRO--GLY-566 DEL.			
RX MEDLINE=99357778; PubMed=10428822;			
RA Passalita P., Pihlajamaa T., Annunen S., Brewton R.G., Wood B.M.,			
RA Johnson C.C., Liu J., Gong Y., Warman M.B., Prockop D.J., Mayne R.,			
RA Ala-Kokko L.;			
RT "Complete sequence of the 23-kilobase human COL9A3 gene. Detection of			
RT Gly-X-Y triplet deletions that represent neutral variants.";			
RL J. Biol. Chem. 274:22469-22475 (1999).			
RN [3]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=21638749; PubMed=11780052;			
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,			
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,			
RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,			
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,			
RA Buck D., Burrill W.D., Butler A.P., Carter C., Carter N.,			
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,			
RA Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,			





```

RESULT 2
ID YHL1_EBV          STANDARD;          PRT;          660 AA.
AC P03181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLFI protein.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OC NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=60871149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211 (1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01555; -; NOT_ANNOTATED_CDS.
DR PIR; A03742; Q0B83.
KW Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86D1D67A37152A2 CRC64;

Alignment Scores:
Pred. No.: 0.00699 Length: 660
Score: 142.50 Matches: 92
Percent Similarity: 32.43% Conservative: 16
Best Local Similarity: 27.63% Mismatches: 125
Query Match: 9.75% Indels: 100
DB: 1 Gaps: 15

US-09-931-704-4 (1-819) x YHL1_EBV (1-660)
QY 806 TGGGCTGAGGTGGGGGTTAAGGTCAGAGGTCAGAAACCATGTGCTCCCAAGTGCA 747
Db 159 TTPArgArgSerGlyAla-----GlnArgGlyHisProProGlyAla 174
QY 746 GGGTGACTGAAGCTGCTGAGGCTGCATCTTCTTAAAGCGGTTGAAGTCTTGGCTG 687
Db 175 GlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThr 194
QY 686 AACGCCATA-----GCCAGGCTGCAGCTCTTTCAGCAGCCAGCAAGTCATCCA 639
Db 195 ProAlaAlaProGlyProGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisPro 214
QY 638 TCTTCTGGAGGAAGTCACCTGTGG-----CAGGGCCAG 606
Db 215 GluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGln 234
QY 605 GGGCCAGGCTGGCTCAGTCCCTGGCAGAGGCTGGGCGAGTGGTAGCAACGCTGCCA 546
Db 235 GluProArgLeuProGlnAspLeuAlaAla----- 245
QY 545 TGACACCTGCAATGTGCTGCCACAGGC----- 519
Db 246 -----GlnArgCysProIadGlyProProThrArgSerGlyAlaAlaGln 262

```

- [3] RP SEQUENCE OF 85-1685 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=90337990; PubMed=2380186;  
RA Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;  
RT "Complete primary structure of the triple-helical region and the  
RT carboxyl-terminal domain of a new type IV collagen chain, alpha  
5(IV)";  
RL J. Biol. Chem. 265:13758-13766(1990).  
RN [4]  
RP SEQUENCE OF 924-1685 FROM N.A.  
RX MEDLINE=91169491; PubMed=2004755;  
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
RT "Characterization of the 3' half of the human type IV collagen alpha  
5 gene that is affected in the Alport syndrome";  
RL Genomics 9:1-9(1991).  
RN [5]  
RP SEQUENCE OF 914-1685 FROM N.A.  
RX MEDLINE=90160375; PubMed=1689491;  
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,  
RA Tryggvason K.;  
RT "Identification of a distinct type IV collagen alpha chain with  
RT restricted kidney distribution and assignment of its gene to the  
RT locus of X chromosome-linked Alport syndrome";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
RN [6]  
RP SEQUENCE OF 1442-1471 FROM N.A.  
RX MEDLINE=90252791; PubMed=2339699;  
RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D.,  
RA Sheer D., Solomon E., Pihlajaniemi T.;  
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene  
RT to the region of the X chromosome containing the Alport syndrome  
RT locus";  
RL Am. J. Hum. Genet. 46:1024-1033(1990).  
RN [7]  
RP SEQUENCE OF 1-20 FROM N.A.  
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,  
RA Marynen P.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).  
RX MEDLINE=94133540; PubMed=8301933;  
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,  
RA Cassiman J.-J., Marynen P.;  
RT "Differential splicing of COL4A5 mRNA in kidney and white blood  
RT cells: a complex mutation in the COL4A5 gene of an Alport patient  
RT deletes the NCI domain";  
RL Kidney Int. 44:1316-1321(1993).  
RN [9]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97338662; PubMed=9195222;  
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
RL "The clinical spectrum of type IV collagen mutations";  
RN Hum. Mutat. 9:477-499(1997).  
RN [10]  
RP VARIANT AS SER-1564.  
RX MEDLINE=91169492; PubMed=1672282;  
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,  
RA Tryggvason K.;  
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a  
RT conserved cysteine to serine in Alport syndrome";  
RL Genomics 9:10-18(1991).  
RN [11]  
RP VARIANT AS ARG-325.  
RX MEDLINE=92303559; PubMed=1376965;  
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,  
RA Tryggvason K., Gubler M.-C., Antignac C.;  
RT "Substitution of arginine for glycine 325 in the collagen alpha 5  
RT (IV) chain associated with X-linked Alport syndrome: characterization  
RT of the mutation by direct sequencing of PCR-amplified lymphoblast  
RT cDNA fragments";  
RL Am. J. Hum. Genet. 51:135-142(1992).  
RN [12]  
RP VARIANT AS GLU-325.  
RX MEDLINE=93244772; PubMed=1363780;  
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,  
RA Rizzoni G.F., de Marchi M.;  
RT "De novo mutation in the COL4A5 gene converting glycine 325 to  
RT glutamic acid in Alport syndrome";  
RL Hum. Mol. Genet. 1:127-129(1992).  
RN [13]  
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
RX MEDLINE=94010948; PubMed=8406498;  
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,  
RA Tryggvason K., Haggma-Schouten W.A.G., Roodvoets A.P., Rascher W.,  
RA van Oost B.A., Smeets H.J.M.;  
RT "Identification of four novel mutations in the COL4A5 gene of  
RT patients with Alport syndrome";  
RL Genomics 17:485-489(1993).  
RN [14]  
RP VARIANTS AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872  
AND C-1241.  
RX MEDLINE=95322976; PubMed=7599631;  
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;  
RT "Detection of 12 novel mutations in the collagenous domain of the  
RT COL4A5 gene in Alport syndrome patients";  
RL Hum. Mutat. 5:197-204(1995).  
RN [15]  
RP VARIANT AS ARG-1649.  
RX MEDLINE=96213750; PubMed=8651292;  
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
RA Denison J.C., Fain P.R., Gregory M.C.;  
RT "A mutation causing Alport syndrome with tardive hearing loss is  
RT common in the western United States";  
RL Am. J. Hum. Genet. 58:1157-1165(1996).  
RN [16]  
RP VARIANTS AS  
RX MEDLINE=96213754; PubMed=8651296;  
RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,  
RA Turco A.E., Hetsikari N., Zhou J., Gusmano R., Massella L., Banfi G.,  
RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,  
RA Savi M., Ballabio A., de Marchi M.;  
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
RT exons of the COL4A5 gene";  
RL Am. J. Hum. Genet. 58:1192-1204(1996).  
RN [17]  
RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND  
MET-1428.  
RX MEDLINE=97094179; PubMed=8940267;  
RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacasier D.,  
RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,  
RA Gubler M.-C., Antignac C.;  
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
RT syndrome";  
RL Am. J. Hum. Genet. 59:1221-1232(1996).  
RN [18]  
RP VARIANT AS ASP-1498.  
RX MEDLINE=96233932; PubMed=8829632;  
RA Tverskaya S., Bobryniina V., Tsalykova F., Ignatova M.,  
RA Krasnopolskaya X., Evgrafov O.;  
RT "Substitution of A1498D in noncollagen domain of alpha 5(IV) collagen  
RT chain associated with adult-onset X-linked Alport syndrome";  
RL Hum. Mutat. 7:149-150(1996).  
RN [19]  
RP VARIANT AS GLN-1677.  
RX MEDLINE=97295089; PubMed=9150741;  
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
RT "Common ancestry of three Ashkenazi-American families with Alport  
RT syndrome and COL4A5 R1677Q";  
RL Hum. Genet. 99:681-684(1997).  
RN [20]  
RP VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.  
RX MEDLINE=98112435; PubMed=9452056;  
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,  
RA Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;



Qy	496	-----CCACTTCTGTACCAAGCCTCCAGAGCGCCCTGCTGGGCGAGCATTCGAGG	540
Db	1017	GlyPheAenglyLeuProGlyglulysGlyglulProgllyProAlaAlaargaspGlyGlu	1036
Qy	541	TGTCAATGCGCAGCGTCTGGCTACCCACTGCCAGCGCTCTGCCAGGAGACTGAG	597
Db	1037	--lysGlyGluProGlyLeuProGlyGlnProGly-LeuArgGlyProGlnGlyProPr	1055
Qy	598	CTGGGCCCTGGCCCTGCCACAGTGCATTCTCCAGAAGATGGATGACTTCTGGCGTGCT	657
Db	1055	oglyLeuProGly-LeuProGlyLeuLysGlyAspGlyGly	1068
Qy	658	GAAGAGCTGCAGACTGGCTGGCTGGCTTCAGCCAGGACTTCAACCGGCTTAAAGAAGAA	717
Db	1069	-----GlnProGlytyrGly-----AlaProGlyLeuMetGlyG	1080
Qy	718	GATGCAGCCTCCAGCAGCTTCAGTCACCCCTGCCTTGGAGGCACATGGTTTCTGACCTCT	777
Db	1080	luLysGlyLeuProGlyLeuProGlyLys-----	1089
Qy	778	GACCTTTAACCCCCACACTCCAGGCCCA	806
Db	1090	--ProGlyArgProGlyAlaProGlyPro	1098
		RESULT 5	
ID	SPD2_NBPCL	STANDARD;	PRT; 627 AA.
AC	P46804;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Spidroin 2 (dragline silk fibroin 2) (Fragment).		
OC	Nephila clavipes (Orb spider).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;		
OC	Araneomorphae; Entelegynae; Araneioidea; Tetragnathidae; Nephila.		
OX	NCBI_TaxID=6915;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92406876; PubMed=1527052;		
RA	Hinman M.B., Lewis R.V.;		
RT	"Isolation of a clone encoding a second dragline silk fibroin.		
RT	Nephila clavipes dragline silk is a two-protein fiber."		
RL	J. Biol. Chem. 267:19320-19324 (1992).		
CC	!- FUNCTION: Spiders major ampullate silk possesses unique		
CC	characteristics of strength and elasticity. Fibroin consists of		
CC	pseudocrystalline regions of antiparallel beta-sheet interspersed		
CC	with elastic amorphous segments.		
CC	!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.		
CC	!- SUBCELLULAR LOCATION: Extracellular.		
CC	!- DOMAIN: Highly repetitive protein characterized by regions of		
CC	polyalanine and glycine-rich repeating units.		
CC	!- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstat-		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for commer-		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/annou">http://www.isb-sib.ch/annou</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; M92913; AAA29381.1; --.		
KW	Silk; Repeat.		
FT	NON_TER 1 1		
FT	DOMAIN 1 530 15 APPROXIMATE TANDEM REPEATS.		
FT	REPEAT 1 36 1.		
FT	REPEAT 37 79 2.		
FT	REPEAT 80 121 3.		
FT	REPEAT 122 172 4.		
FT	REPEAT 173 213 5.		
FT	REPEAT 214 252 6.		
FT	REPEAT 253 283 7.		

```

FT REPEAT 284 317 8.
FT REPEAT 318 359 9.
FT REPEAT 360 391 10.
FT REPEAT 392 428 11.
FT REPEAT 429 464 12.
FT REPEAT 465 488 13.
FT REPEAT 489 515 14.
FT REPEAT 516 530 15.
SQ SEQUENCE 627 AA; 54184 MW; CB9B63779B2C594B CRC64;

Alignment Scores:
Pred. No.: 0.0276 Length: 627
Score: 134.00 Matches: 73
Percent Similarity: 36.44% Conservative: 13
Best Local Similarity: 30.93% Mismatches: 79
Query Match: 9.17% Indels: 71
DB: Gaps: 12

US-09-931-704-4 (1-819) x SPD2_NEPCL (1-627)
QY 671 TCTGAGCTCCTTCAGCAGCCAGAGTCATCCATCTTCTGGAGGAAGTCACTGTGGCAG 612
D 245 SerAlaAlaAlaAlaAlaAlaGlyProGlyGlnGlyProGlyGlyProGlyPro 264
QY 611 GGC---CAGGGCCAGGCTGCTAGTCTCTGGCAGAGGCTGGGCGAGTGGTAGCAA 555
D 265 GlyGlnGlnGlyProSerGlyProGlySerAlaAlaAlaAlaAlaAlaAlaAla 282
QY 554 GCGTCGCGATGACACCTGCAATGCTGCCAGCAGGCGCTGAGGCTGGTACAGAGTGGG 495
D 283 ---AlaGlyProGlyGlyProGlyProGlyProGlyProGlyProGlyProGly 289
QY 494 CCAGGCTACGTCGGAGTTCAGCTGTGGCAGCCCTGAGGCTGGTACAGAGTGGG 435
D 290 ProGlyGlnGlnGly-----ProGlyGlyProGly----- 299
QY 434 GAGGTGACTGTACGCTCATAGTCTTGGTTCAGCCGACGCTGTCTATGAGGCTGCC 375
D 300 ProGlyGlnGlnGlyProSerGlyAlaGlySerAlaAlaAlaAlaAlaAlaAla 319
QY 374 ACACITCCAA-----GTTGACCGTGGCTGGCGAGGTTCTGCCCCAGTCGAGGAG 321
D 319 oGlyGlnGlnGlyLeuGlyGlyProGlyProGlyGln----- 331
QY 320 GATTCAGAGTCAGCTGTTGAAGGGGGCCCGCAGGTAGTTCAGGTAGTCCAGCTAAGC 261
D 332 -----GlnGly--ProGlyGlyProGlyProGlyGlnGly 343
QY 260 TCCGAGTGTATGCTCCAGTAGC-----GGGTGAGGTGTCAT 225
D 344 ProGlyGlyTyroGlyProGlySerAlaSerAlaAlaAlaAlaAlaAlaAlaAla 363
QY 224 AGGTTTCTGGATGGAGGGCCAGGSCCTGGATCTCTGTCGATTAAGAGTGGCAGTG 165
D 364 GlnGlyProGlyGlyTyroGlyProGly----- 372
QY 164 CAGGAGGTGCCACAGCCGTCATAGGACAGCTAACATCCCGCAGCTCCCTGCTC 105
D 373 GlnGlnGlyProSerGlyPro-----GlySerAlaSerAlaAlaAlaAlaAlaAla 390
QY 104 GGAGTCCATGGGCTGGGGGGCGGGCGCGCGCGCGGCGG-----GCTCTCTCCCA 54
D 391 AlaGlyProGlyGlyTyroGlyProGlyGlnGlnGlyProGlyGlyTyroGlyGln 410
QY 53 GAGGCT---GGCGAGTGGAGGGCGAGCGCGGCTCCGCG 15
D 411 GlnGlyProSerGlyProGlySerAlaSerAlaAlaAlaAlaAlaAlaAlaAla 424

RESULT 6
T2D3 HUMAN STANDARD; PRT; 1083 AA.
AC O00258; Q9B721; Q9B742; Q9B740;
DT 15-JUL-1998 (Rel. 36, Created)

```

```

DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)
DE (TAFII135) (TAFII-120) (TAFII130).
GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97336072; PubMed=9192867;
RA Mengus G., May M., Carre L., Chambon P., Davidson I.;
RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
RT of the retinoic acid, vitamin D3, and thyroid hormone receptors in
RT mammalian cells";
RL Genes Dev. 11:1381-1395(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blake S.E., Bridgman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gillman R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehaialho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McComachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhan R., Sims S.,
RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=97098442; PubMed=8942982;
RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RT "Molecular cloning and analysis of two subunits of the human TFIID
RT complex: hTAFII30 and hTAFII100.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT
CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
CC BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y11354; CAA72189.1; -.
CC EMBL; AL137077; CAC36006.1; -.
DR

```

```

DR EMBL; AL109911; CAC22312.2; -.
DR EMBL; U75308; AAC50901.1; -.
DR TRANSFAC; T02328; -.
DR Genew; HGNC:11537; TAF4.
DR MIM; 601796; -.
DR InterPro; IPR003894; TAF_hom.
DR SMART; SM00549; TAFH; 1.
KW Transcription regulation; Nuclear protein.
FT DOMAIN 39 42 POLY-HIS.
FT DOMAIN 52 57 POLY-ALA.
FT DOMAIN 98 101 POLY-GLY.
FT DOMAIN 148 148 POLY-ALA.
FT DOMAIN 268 275 POLY-PRO.
FT DOMAIN 331 337 POLY-ALA.
FT DOMAIN 680 683 POLY-PRO.
FT DOMAIN 808 813 POLY-ALA.
FT DOMAIN 828 831 POLY-ASP.
FT DOMAIN 105 117 PGPPSPRRPLVPA -> GRLLQQRGGRES
FT CONFLICT 136 136 A -> S (IN REF. 2).
FT CONFLICT 185 185 G -> GPG (IN REF. 2).
FT CONFLICT 233 264 MISSING (IN REF. 3).
FT CONFLICT 293 293 P -> L (IN REF. 3).
SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;

Alignment Scores:
Pred. No.: 0.0326 Length: 1083
Score: 133.00 Matches: 81
Percent Similarity: 35.45% Conservative: 25
Best Local Similarity: 27.09% Mismatches: 123
Query Match: 8.88% Indels: 70
DB: 1 Gaps: 14

US-09-931-704-4 (1-819) x T2D3_HUMAN (1-1083)
Qy 19 GGAGCGCGGCTGCGCCTCCCACTCGCCAGCTCTGGAGAGAGAGCGCGCGCGCGG 78
Db 81 GlyAlaProGlyAlaAlaProGluProProProAlaGlyArg--AlaArgProGlyGly 99
Qy 79 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 132
Db 100 GlyGlyProGluArgProGlyProProSerProArgProAlaGlyPro 119
Qy 133 TTGCCTATGACGCGTCTGCGCACCTCTGCGAGTCTTAATCGCACAGGAGA 192
Db 120 AlaProProAlaAlaLysLeuArgProProProGluGlySerAlaGlyAlaCysAla--- 138
Qy 193 TCCAGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 252
Db 139 -----ProValProAlaAlaAla 144
Qy 253 ACT-----CCGCGAGCTAGTGGACCTACCTGAACCTAGTGGCGCGCGCGCTTT 300
Db 145 AlaValAlaAlaGlyProGluProAlaProAlaGlyProAlaLysProAlaGlyProAla 164
Qy 301 CAACGAGCGCTGACTTCAATCTCTCGACTGGG----- 333
Db 165 AlaLeuAlaAlaArgAlaGlyProGlyProGlyProGlyProGlyProGlyPro 184
Qy 334 GCGAGAACTGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAA 387
Db 185 GlyLysProAlaGlyProGlyAlaAlaGlnThrLeuAsnGlySerAlaAlaLeuLeuAsn 204
Qy 388 TCAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 444
Db 205 SerHisHisAlaAlaAlaProAlaAlaValSerLeuValAsnAsnGlyProAlaAlaLeuLeu 224
Qy 445 GCGTGGCGCTCAACCGTCAGCGCTGCCACAGCTGAACCTCCGAGTACGCTGGCGCGCGCTCTGT 504
Db 225 ProLeuProLysProAlaAla-----ProGlyThrValIle 236
Qy 505 TACGAGCTTCCA-----GGCGCTGCTGGCGAGCATTCGACGTTGTCATGGC 549

```

```

Db 237 GlnThrProProPheValGlyAlaAlaProProAlaProAlaProAlaProSerProPro 256
Qy 550 GAGCGTTGGCTACCCAGCTGCCAGCTCTGCTCCAGGAGCTGAGCCAGCGCTGGCGCCCTGG 609
Db 257 AlaAlaProAlaProAlaAlaProAlaAlaProPro-----ProProPro 272
Qy 610 CCTGCCCCACAGTGACTT-----CCTCCA-----GAAGATGGATGACTTCTGGCTGCT 657
Db 273 ProAlaProAlaThrLeuAlaArgProProGlyHisProAlaGlyProProThrAlaAla 292
Qy 658 GAAGGAGCTCCAGACCTGGCTATGGCTTTCAGCCCAAGGACTTCAACCGGCTTAAGAAGA 717
Db 293 ProAlaValProProProAlaAlaAlaAlaGlnAsnGlyGly----- 305
Qy 718 GATGCGACCTCCAGCAGCTTTCAGTCACCTGCGACCTGGAGGACCATGTTTCTGACCTCT 777
Db 306 SerAlaGlyAlaAlaProAlaProAlaProAlaGlyGlyProAlaGlyValSerGly 325
Qy 778 GACCTTTAACCC-----CCACACCTCCAGGCGCGCGCTCA 810
Db 326 GlnProGlyProGlyAlaAlaAlaAlaProAlaProGlyValLysAlaGluSer 344

RESULT 7
CAL4_MOUSE
ID CAL4_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89137932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glangville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen."

```

J. Biol. Chem. 262:8496-8499 (1987).  
 [6]  
 RN PARTIAL SEQUENCE FROM N.A.  
 RP MEDLINE=86196099; PubMed=3009468;  
 RA Sakurai Y., Sullivan M., Yamada Y.;  
 RT "Alpha 1 type IV collagen gene evolved differently from fibrillar  
 collagen genes.";  
 RL J. Biol. Chem. 261:6654-6657 (1986).  
 [7]  
 RN SEQUENCE OF 1-28 FROM N.A.  
 RP MEDLINE=89066738; PubMed=3198626;  
 RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;  
 RT "Head-to-head arrangement of murine type IV collagen genes.";  
 RL J. Biol. Chem. 263:19274-19277 (1988).  
 [8]  
 RN SEQUENCE OF 1-28 FROM N.A.  
 RP MEDLINE=89071759; PubMed=3200851;  
 RA Burdello P.D., Martin G.R., Yamada Y.;  
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a  
 bidirectional promoter and a shared enhancer.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682 (1988).  
 [9]  
 RN SEQUENCE OF 1-129 FROM N.A.  
 RP MEDLINE=88243724; PubMed=3379041;  
 RA Killen P.D., Burdello P., Sakurai Y., Yamada Y.;  
 RT "Structure of the amino-terminal portion of the murine alpha 1(IV)  
 collagen chain and the corresponding region of the gene.";  
 RL J. Biol. Chem. 263:8706-8709 (1988).  
 CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/  
 CC NIDOGEN.  
 CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; J03758; AAA37439.1; -  
 CC EMBL; M23333; AAA51625.1; -  
 CC EMBL; J04594; AAA50292.1; -  
 CC EMBL; X06777; CAA29946.1; -  
 CC EMBL; X02201; CAA26132.1; -  
 CC EMBL; M15832; AAA37340.1; -  
 CC EMBL; M14042; AAA37342.1; -  
 CC EMBL; M12879; AAA37343.1; -  
 CC EMBL; M13024; -; NOT ANNOTATED CDS.  
 CC EMBL; M13025; -; NOT ANNOTATED CDS.  
 CC EMBL; M13026; AAA37344.1; -  
 CC EMBL; M13027; AAA37345.1; -  
 CC EMBL; M13043; AAA37346.1; -  
 CC EMBL; J04448; AAA37437.1; -  
 CC PIR; A33525; CGMS48.  
 CC MGD; MGI:88454; Col4a1.  
 CC InterPro; IPR000087; Collagen.  
 CC InterPro; IPR001442; ProcollagenC4.

DR Pfam; PF01391; Collagen; 23.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD000007; Collagen; 4.  
 DR ProDom; PD003923; ProcollagenC4; 2.  
 DR SMART; SM00111; C4; 2.  
 KW Extracellular matrix; Connective tissue; Basement membrane;  
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 27  
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.  
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).  
 FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).  
 FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).  
 FT DISULFID 1505 1511 BY SIMILARITY.  
 FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).  
 FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).  
 FT DISULFID 1616 1622 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CONFLICT 26 26 A -> P (IN REF. 2).  
 FT CONFLICT 186 186 S -> L (IN REF. 2).  
 FT CONFLICT 319 319 Q -> S (IN REF. 2).  
 FT CONFLICT 369 369 Q -> L (IN REF. 2).  
 FT CONFLICT 403 403 L -> F (IN REF. 2).  
 FT CONFLICT 481 481 P -> L (IN REF. 2).  
 FT CONFLICT 493 493 Q -> H (IN REF. 2).  
 FT CONFLICT 712 712 S -> I (IN REF. 2).  
 FT CONFLICT 813 813 E -> Q (IN REF. 2).  
 FT CONFLICT 982 982 Q -> H (IN REF. 2).  
 FT CONFLICT 1397 1397 V -> S (IN REF. 3).  
 SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058B9 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.0384 Length: 1669  
 Score: 132.00 Matches: 75  
 Percent Similarity: 34.26% Conservative: 11  
 Best Local Similarity: 29.88% Mismatches: 91  
 Query Match: 8.81% Indels: 75  
 DB: 1 Gaps: 15  
 US-09-931-704-4 (1-819) x CA14\_MOUSE (1-1669)  
 QY 13 TTCGCGGAGCGCGGCTGCGCTCCACTCCGCGCAGCCTCTGGGAGAGAGCGCGGCC 72  
 Db 1193 PheProGlyLeuAlaGlySerProGlyIleProGlyValLysGlyGluGlnGlyPheMet 1212  
 QY 73 GCGCGCGCGCGCGCCCGAG-----CCCCAT 96  
 Db 1213 GlyProProGlyProGlnGlyGlnProGlyLeuProGlyThrProGlyHisProValGlu 1232  
 QY 97 GGACCTCCG---AGCAGGGGACTCGTGGGGGATGTTAGCTTCCTATGCAC---GGTGCT 150  
 Db 1233 GlyProLysGlyAspArgGlyProGlnGlnProGlyLeuProGlyHisProGlyPro 1252  
 QY 151 GTGGCACTTCCTGCGAGTGCAGCTCTTAAT---CGCACGAGGATCCAGGCGCC----- 201  
 Db 1253 MetGlyPro-ProGlyPheProGlyIleAsnGlyProLysGlyAspLysGlyAsnGlnG 1272  
 QY 202 -TGGCCCTCCATCCAGAAACCTATGACTCACC----- 237  
 Db 1272 yTrpProGlyAlaProGlyVal-----ProGlyProLysGlyAspProGlyPheGlnG 1290  
 QY 238 -----CTACCTGGAGCATCACTCCGAGCTTAGCTGGGAGCTACC 278  
 Db 1290 yMetProGlyIleGlySerProGlyIleThrGlySerLysGlyAspMetGlyLeuPr 1310  
 QY 279 TGAACCTACTGGGCGCCCTTTCAACGAGCCTGACTTCAATCCTCCTCGACTGGGGGCGAG 338  
 Db 1310 oGlyValProGlyPheGlnGlyGlnLys-----GlyLe 1321  
 QY 339 AAACCTGCTCCCGAGGC---CAGGTCACTTGGAGTGTGGGAAGCCTCAATGACAGGC 395  
 Db 1321 uProGlyLeuGlnGlyValLysGlyAspGlnGlyAsp-----GlnG 1335

QY	396	TGGCGCTGACCCAGAACTATGAGGGGTACAGTCACTCCTCTGTGTTACTTGGCTGGCTCA	455
Db	1335	yValProGlyProlys-----GlyLeuGlnGlyProPro-----GlyProProGly	1350
QY	456	ACCGTCAGGCTGCCACAGCTGAACCTCCGACGTAGCTGGCCCACTT-----	501
Db	1350	yProTyAspValIleIleIleGlyGluProGlyLeuProGlyProGluGlyProProGlyLe	1370
QY	502	-----CTGTACCACTCCAGGCTCTGCTGGGAGCATTTGACGTGTCATGGCAGCG	554
Db	1370	uLysGlyLeuGlnGlyProProGlyProIleGlyGln-Gln-----GlyValThrGlyser	1389
QY	555	TGGGTACCACTGCCAGCTCTGCCAGG-----	586
Db	1389	alGlyLeuProGlyProProGlyValProGlyPheAspGlyAlaProGlyGlnLysGlyG	1409
QY	587	--ACTGAGCGAGCTGGGCGCTGGCCCT 613	
Db	1409	luThrGlyProPheGlyProProGlyPro 1418	
RESULT 8			
ID	YQ35 CAEEL	STANDARD; PRT; 317 AA.	
AC	Q09456;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	01-NOV-1997 (Rel. 35, Last annotation update)		
DE	Putative cuticle collagen C09G5.5.		
GN	C09G5.5.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Bristol N2;		
RA	Palmer S.;		
RL	Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.		
CC	!- FUNCTION; NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE		
CC	PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A		
CC	BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).		
CC	!- SUBUNIT; COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE		
CC	CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-		
CC	LINKS (BY SIMILARITY).		
CC	!- SIMILARITY; BELONGS TO THE CUTICULAR COLLAGEN FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
DR	EMBL; Z46791; CAA86758.1; --		
DR	WormBep; C09G5.5; CE01485.		
DR	InterPro; IPR002486; Col_cuticle_N.		
DR	InterPro; IPR000087; Collagen.		
DR	Pfam; PF01391; Collagen; 3.		
DR	Pfam; PF01484; Col_cuticle_N; 1.		
KW	Hypothetical protein; Cuticle; Connective tissue; Repeat;		
FT	DOMAIN 92 124 TRIPLE-HELICAL REGION.		
FT	DOMAIN 137 199 TRIPLE-HELICAL REGION.		
FT	DOMAIN 202 264 TRIPLE-HELICAL REGION.		
SQ	SEQUENCE 317 AA; 31283 MW; 685DCF24612707BB CRC64;		
Alignment Scores:			
'pred. No.:	0.0446	Length:	317
Score:	131.00	Matches:	63
Percent Similarity:	34.08%	Conservative:	13
Best Local Similarity:	28.25%	Mismatches:	68

Query Match:	8.74%	Indels:	79
DB:	1	Gaps:	13
US-09-931-704-4 (1-819) x YQ35 CAEEL (1-317)			
QY	19	GGAGCGCGCTGCGCTCCCTCCACTCCCGCAGCCTCTGGG-----	57
Db	110	GlyProAlaGlyLysProGlyGlnProGlyValAlaGlyProAlaHisHisGlnGlnGln	129
QY	58	-----AGAGGAGGCC-----GGCGCGCGCGCGCGCGC	84
Db	130	GlucyIleIleLysCysProGlnGlyAlaProGlyProAlaGlyAlaProGlyAsnProGly	149
QY	85	CCCCAG---CCCATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCTTGCTATG	141
Db	150	ProGlnGlyProAsnGlyAsnPro-----GlyAlaProAla	161
QY	142	CAGGTGCTGTGGCACCTCCCTGCGAGTCCAGCTCTTAATCGCACAGGAGATCCAGGCC	201
Db	162	HisGlyGlyGlyGlnGlyPro-----Pro	169
QY	202	TGGCGCTCCATCCAGAAACCTATGACCTCACCGCTACCTGGAGCATCAACTCCGCGAG	261
Db	170	GlyPro-----ProGlyProAlaGlyAspAlaGlySerProGly	182
QY	262	CTTAGCTGGGACCTACCTGAACCTACCTGGGGCCCTTTTCAACGAGCCTGACTTCAATCC	321
Db	183	GlnAla---GlyAlaProGlyAsnProGlyArgPro-----	193
QY	322	TCCTCGACTGGGGGAGAACTCTGCCAGGCGCCAGCTTCACTTGGAGTGTGGCGAAG	381
Db	194	-----GlyGlnSerGlyGlnArgSerArgGlyLeuProGlyProSerGlyArg	209
QY	382	CCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGCGGTACAGTCACCTCTGTGTGA	441
Db	210	Pro-----GlyProGlnGlyProPro-----	216
QY	442	CTTGGCTGGCTCAACCTCAGCTGAGGTGCCACAGCTCACTCCGACGTAGCTGGCCACIT	501
Db	217	---GlyAlaProGlyGlnProGlySerGlySer---ThrProGlyProAlaGlyPro---	233
QY	502	CTGTACCACTCCAGGCGCTGCTGGGAG-----CATTCAGGTGTTCATGGCGACGCT	555
Db	234	---ProGlyProProGlyProAsnGlyGlnProGlyHis-ProGlyGlnAspGlyGlnPr	252
QY	556	TGGTACCACTGCGCCAGCCTCTCCAGGAGTACAGCCAGCTGGGGCCCTGGCCCTGC	615
Db	252	oGlyAlaProGlyAsnAspGlyAlaProGlySerAspAlaAlaTyrCysProCysProAl	272
QY	616	CAACAGT 622	
Db	272	aArgSer 274	
RESULT 9			
ID	CA54 CANFA	STANDARD; PRT; 754 AA.	
AC	Q28247;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	15-JUL-1998 (Rel. 36, Last annotation update)		
DE	Collagen alpha 5(IV) chain (Fragment).		
GN	COL4A5.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Samoyed; TISSUE=Kidney;		
RC	MEDLINE=94224868; PubMed=8171024;		
RA	Zheng K.; Thorne P.S.; Marrano P.; Bauml R.; McInnes R.R.;		
RT	"Canine X chromosome-linked hereditary nephritis: a genetic model for		
RT	human X-linked hereditary nephritis resulting from a single base		



mutation in the gene encoding the alpha 5 chain of collagen type IV.";

RT Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).

CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/NIDOGEN.

CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.

CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).

CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.

CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.

CC -1- DISEASE: A DEFECT IN COL4A5 HAS BEEN FOUND TO BE THE CAUSE OF CANINE X-LINKED HEREDITARY NEPHRITIS (HN), A DISEASE SIMILAR TO THAT IN HUMANS (ALSO REFERRED TO AS ALPORT SYNDROME) CHARACTERIZED BY PROGRESSIVE RENAL FAILURE AND NEUROSENSORY DEAFNESS.

CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U07888; AAB60258.1; --

DR InterPro; IPR000087; Collagen.

DR InterPro; IPR001442; Procollagnc4.

DR Pfam; PF01391; Collagen; 8.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD003923; Procollagnc4; 2.

DR SMART; SM00111; C4; 2.

DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.

KW NON TER 1

FT DOMAIN <1 530 TRIPLE-HELICAL REGION.

FT DOMAIN 531 >754 NONHELICAL REGION (NC1).

FT DISULFID 552 643 OR 640 (BY SIMILARITY).

FT DISULFID 585 640 OR 643 (BY SIMILARITY).

FT DISULFID 597 603 BY SIMILARITY.

FT DISULFID 662 ? OR 754 (BY SIMILARITY).

FT DISULFID 696 754 BY SIMILARITY.

FT DISULFID 708 714 BY SIMILARITY.

FT NON\_TER 754 754

SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;

Alignment Scores:

Pred. No.:	0.0449	Length:	754
Score:	131.00	Matches:	62
Percent Similarity:	34.70%	Conservative:	14
Best Local Similarity:	28.31%	Mismatches:	76
Query Match:	8.74%	Indels:	67
DB:	1	Gaps:	13

US-09-931-704-4 (1-819) x CA54\_CANFA (1-754)

QY 19 GGAGCGCGGCTCGCCCTCCACTCCGCCAGCCCTCTGGGAGAGGAGCGCGCCGCCGCG 78

Db 180 GlyLeuProGlyLeuProGlyThrProGlyAlaLysGlyGlnProGlyLeuProGlyPhe 199

QY 79 CCGGCG-----CCCGAGCCC-----CATGGACCTCCGAGCAGG---GGA 114

FT REPEAT 399 523 3.  
FT REPEAT 524 648 4.  
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

## Alignment Scores:

Pred. No.: 0.0527 Length: 660  
Score: 130.00 Matches: 91  
Percent Similarity: 31.27% Conservative: 15  
Best Local Similarity: 26.84% Mismatches: 97  
Query Match: 8.68% Indels: 136  
DB: 1 Gaps: 20

US-09-931-704-4 (1-819) x YHLL\_EBV (1-660)

QY 19 GGAGCGGGCTCGCCCTCCCACTCGCCAGCTCTGGGAGAGAGCGCGCGCGCGG 78  
Db 173 GlyAlaGlyGlnArg---ProSerGlyProThrGlyGlyArgProAlaAlaProGlyAla 191  
QY 79 CCGGCG---CCCGAGCGCCCATGGA----- 99  
Db 192 ProGlyThrProAlaAlaProGlyProGlyGlyGlyAlaAlaValProSerGlyAlaThr 211  
QY 100 -----CCTCCG 105  
Db 212 ProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProPro 231  
QY 106 AGCAGGGGACTCGTGGGATGTAGCTTGCCTATGACGGTCTGTGGCACCTCCCTGC 165  
Db 232 GluArg-----GlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCys 248  
QY 166 -----AGTCCAGCTCTTAATCGCACGAGAGA--- 192  
Db 249 ProAlaGlyProProThrArgSerGlyAlaAlaGlnArgThrHisArgArgPro 268  
QY 193 -----TCCAGGCGCTGGCCCTCCATCCA 216  
Db 269 ProGlyCysProArgSerAlaArgAsnProGlyCysProArgThrProArgArgSer 288  
QY 217 GAAACCTATGACCTCACCGCTACTCGGAGCA-----TCAACTCCGCGAGCTT 264  
Db 289 GlyAlaGlnArgGlyHisProProGlyAlaGlyGlnArgProSerGlyPro----- 306  
QY 265 AGCTGGGACTTACCTGAACCTACCTGGGGCCCCCTTTCAACGAGCCTGAACCTCCTCC 324  
Db 307 ThrGlyGlyArgProAlaAlaProGlyAlaPro----- 317  
QY 325 TCGACTGGGGGAGAACTCTGCCAGGCGCCACGTCACCTTGAAGTGTGGGAGCCT 384  
Db 318 -----GlyThrProAlaAlaProGlyProGly-----GlyGlyAlaAlaValPro 332  
QY 385 CAATGACAGGCTCGGCTGACCCAGAACTA---TGAGGCGTACAGTCACCTCCTGTGTTA 441  
Db 333 Ser---GlyAlaThrProHisProGluArgGlySerGlyProAlaAspProProAlaAla 351  
QY 442 CTTCGTGGCCT-----CAACCGTCAGGCTGCCACAG---CTGAACCTCCGAGC 486  
Db 352 AlaArgLeuProProGluArgGlnGluPro--ArgLeuProGlnAspLeuAlaAlaAlaG 371  
QY 487 TAGCCTGGCCACTCTGTACACGCTCCAGGCGCTCTGGGAGCATTGCGAGTGTAT 546  
Db 371 lnArgCysProAlaGlyProProThrArgSer---GlyAlaAlaGlnArgThrH 390  
QY 547 GGGAGCTGGCTACCTACCTGCCCCAGCCTCTGCCAGGAGCTGAGCCAGCCTGGGCCCC 606  
Db 390 iArgArg---ProProGlyCysPro----- 397  
QY 607 TGCCCTGCCACAGTGAACCTTCCAGAGATGGATGATCTTCTGGTGTGTAAGAGACT 666  
Db 398 -----ArgSerAlaArgAsnProGlyCysProArgThrT 409  
QY 667 GCAGA-----CCTGGCTATGGCTTCAGCCAA 693  
Db 409 rpArgArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaGlyGlnArgProS 429

QY 694 GGACTTCAACCGGCTTAAGAGAAGATGCAGCCTTCCAGAGCTTCAGTCAACCTGCCTT 753  
Db 429 erGlyProThrGlyGlyArgPro-----AlaAlaProGlyAla- 441

QY 754 GGAGGCACATGGTTCTGACCTTAACCCACACACCTCCAGGCCCA 806  
Db 442 -----ProGlyThrProAlaAlaProGlyPro 450

## RESULT 11

CCD2\_CAEEL STANDARD; PRT; 360 AA.  
ID CCD2\_CAEEL  
AC F35799; Q22477;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cuticle collagen dpy-2 precursor.  
GN dpy-2 OR T1484.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=94060446; PubMed=8241567;  
RA Levy A.D., Yang J., Kramer J.M.;  
RT "Molecular and genetic analyses of the Caenorhabditis elegans dpy-2  
RT and dpy-10 collagen genes: a variety of molecular alterations affect  
RT organismal morphology.";   
RL Mol. Biol. Cell 4:803-817 (1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Pavello T.;  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
CC LINKS.  
CC -1- DISEASE: MUTATIONS IN DPY-2 AFFECTS THE BODY SHAPE.  
CC -1- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL; L12706; AAA17398.2; --  
CC EMBL; U50191; AAK31557.1; --  
CC WormPep; T1484.6; CE04953.  
CC InterPro; IPR002486; Col\_cuticle\_N.  
CC InterPro; IPR000087; Collagen.  
CC Pfam; PF01391; Collagen; 2.  
CC Pfam; PF01484; Col\_cuticle\_N; 1.  
CC Cuticle; Connective tissue; Repeat; Multigene family; Collagen;  
CC Signal.  
CC  
CC SIGNAL  
CC CHAIN 1 ?  
CC DOMAIN 123 360  
CC DOMAIN 152  
CC DOMAIN 174 230  
CC DOMAIN 238 303  
CC VARIANT 129 129  
CC VARIANT 183 183  
CC VARIANT 247 247  
CC VARIANT 253 253  
CC CONFLICT 355 360  
CC  
CC POTENTIAL.  
CC CUTICLE COLLAGEN DPY-2.  
CC TRIPLE-HELICAL REGION.  
CC TRIPLE-HELICAL REGION.  
CC TRIPLE-HELICAL REGION.  
CC G -> E (IN DPY2 (SC38)).  
CC G -> R (IN DPY2 (E8)).  
CC G -> R (IN DPY2 (E489)).  
CC G -> R (IN DPY2 (Q292)).  
CC FRIRKW -> DQYNGESVSNHKNNGYVHLRFTQ (IN  
CC REF. 2).

'SQ SEQUENCE 360 AA; 36738 MW; B827960838A22578 CRC64;

## Alignment Scores:

Pred. No.: 0.0569 Length: 360  
Score: 129.50 Matches: 60  
Percent Similarity: 31.67% Conservativity: 10  
Best Local Similarity: 27.15% Mismatches: 69  
Query Match: 8.64% Indels: 82  
DB: 1 Gaps: 11

US-09-931-704-4 (1-819) x CCD2\_CABEL (1-360)

QY 16 GCGGAGCGCGCTCGCCCTCCACTCCGCCAGCTCTGGAGAGAGCGGCCCGGC 75  
DB 128 SGGGASpSerGlyLeuProAlaLeuProGlyAlaProGlyAlaProGly 147  
QY 76 CGGCG 135  
DB 148 ArgProGlyThrProAlaSer-----Cysile 158  
QY 136 CCTATGACAGGTCTGTGGCACTCTCCCTGAGTGCAGCTCTTAATCGCAGGAGATCC 195  
DB 159 ProGluArgValPheGluProProProCys----- 168  
QY 196 AGGCGTGGCGCTCCATCCAGAAACCTATGACCTCACCGGTACCTGGAGCATCACT 255  
DB 169 LeuProCysProGlnGlyProArgGlyValProGlyHis----- 181  
QY 256 CGCGAGCTTAGTGGGACCTACCTGAACCTACCTGGGCGCGCGCGCGCGCGCG 315  
DB 182 -----ProGlyPheProGlyAspProGlyGlu----- 190  
QY 316 CAATCTCTCGACTGGGCGAGAACTCTGCCAGGCGCCAGCTCACTTGAAGTGTG 375  
DB 191 -----TyrGlyIleGlyArg-----ProGlySerAspGlyMetPro 203  
QY 376 GCGAGCGCTCAATGACAGGTGGCGTGGCGAGAACTATGAGCGGTACAGTCACTCCT 435  
DB 204 GlyLysPro-----GlyAspProGlyLeuAlaGlyProIleGlyProPro 218  
QY 436 -----GTGTACTTGGC 447  
DB 219 GlyGluSerGlyProIleGlyAspLysGlyArgThrProGluAlaHisValIleProGly 238  
QY 448 TGGCTCAACCTCAGCTGAGTGGCGAGCACTGAGTGGCGAGTGGCGAGTGGCGAG 507  
DB 239 ProProGlyAspSerGlyLeu-----ProGlyPro-----Tip 249  
QY 508 CAGCGCTCAGGCGCTGCGGCGAG-----CATTCAGGTGTCTATGGCGAC 552  
DB 250 GlyProProGlySerAlaGlyMetProGlyGluAspGlyTyr-AlaGlyThrProGlyG1 269  
QY 553 GCTTGGTACCCAGTCCCGCGAGCTCTGCGAGGAGTGGCGAGTGGCGAGTGGCGAG 612  
DB 269 uLysGlyTrpProGlyProProGlyAlaProGly-----ProGlyGlyMetProGlyPr 287  
QY 613 T 613  
DB 287 0 287

## RESULT 12

ID \_CC07\_CABEL STANDARD; PRT; 316 AA.  
AC P18832; Q93210;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cuticle collagen 7 precursor.  
GN COL-7 OR C15A11.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;

RN SEQUENCE FROM N.A.  
RP STRAIN=Bristol N2;  
RA Gardner A.E.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.  
[2]  
RN SEQUENCE OF 1-181 FROM N.A.  
RP STRAIN=Bristol N2;  
RX MEDLINE=89326131; PubMed=2753356;  
RA Cox G.N., Fields C., Kramer J.M., Rosenzweig B., Hirsh D.;  
RT "Sequence comparisons of developmentally regulated collagen genes of  
RL Caenorhabditis elegans";  
CC Gene 76:331-344 (1989).  
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
CC LINKS.  
CC -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z79694; CAB01961.1; -;  
CC EMBL; M25478; AAA27992.1; -;  
CC Wormpep; C15A11.5; CE08173.  
CC PIR; PS0036; PS0036.  
CC InterPro; IPR002486; Col\_cuticle\_N.  
CC InterPro; IPR000087; Collagen.  
CC Pfam; PF01391; Collagen; 3.  
CC Pfam; PF01484; Col\_cuticle\_N; 1.  
CC KwdCuticle; Connective tissue; Repeat; Multigene family; Collagen;  
KW Signal.  
FT CHAIN 1 34 POTENTIAL.  
FT DOMAIN 35 316 CUTICLE COLLAGEN 7.  
FT DOMAIN 94 126 TRIPLE-HELICAL REGION.  
FT DOMAIN 139 198 TRIPLE-HELICAL REGION.  
FT DOMAIN 204 263 TRIPLE-HELICAL REGION.  
FT CONFLICT 48 60 MISSING (IN REF. 2).  
FT CONFLICT 165 165 G -> P (IN REF. 2).  
SQ SEQUENCE 316 AA; 31240 MW; 7C311F63FCBD702B CRC64;  
  
Alignment Scores:  
Pred. No.: 0.0725 Length: 316  
Score: 128.00 Matches: 61  
Percent Similarity: 32.37% Conservativity: 17  
Best Local Similarity: 25.31% Mismatches: 72  
Query Match: 8.54% Indels: 91  
DB: 1 Gaps: 12

US-09-931-704-4 (1-819) x CC07\_CABEL (1-316)

QY 28 GCTCGCCCTCCACTCCGCCAGCTCT-----GGG 57  
DB 93 AlaGlyProProGlyProProGlyAlaSerGlyAspLysGlyHisAspGlyGlnProGly 112  
QY 58 AGAGGAGCG 84  
DB 113 GlnAlaGlyLysProGlyGlnProGlyValAlaGlyProSerHisGlnLysGlnGlu 132  
QY 85 -----CCCCAGCCCATGACCTCCGAGC----- 108  
DB 133 CysIleLysCysProGlnGlyLeuProGlyProAlaGlyValProGlyGlnProGlyPro 152  
QY 109 AGGGAGCTCTGGGGATGTTAGTTCCTATGACGCGTGTGCGGACCTCCCTGCGAGT 168  
DB 153 LysGlyProAsnGlyAsnProGlyAlaProAlaGlnGlyGlyGlnGlyPro----- 170

Qy 169 GCCAGCTCTTAATCGACAGGAGATCCAGGCGCTGCGCCCTCATCCAGAAACCTATGA 228  
 Db 171 -----ProGlyPro----- 173  
 Qy 229 COTACCCGCTACTGGAGCATCAATCCGCGAGCTTAGCTGGGACCTACTGAACTACCT 288  
 Db 174 ProGlyProAlaGlySerAlaGlySerProGlyGlnAla---GlyAlaProGlyAsnPro 192  
 Qy 289 GGGGCCCCCTTTCAACGAGCTGACTTCAATCTCTCGACTGGGGGCGAGAACTCTGCC 348  
 Db 193 GlySerPro-----GlyLySerGlyGln 200  
 Qy 349 CAGGCGCCAGCTCACTTGGAACTGTGGGAGCCCTCAATGACAGGCTGGGCTGACCCA 408  
 Db 201 ArgGlyArgGlyLeuProGlyProSerGlyAlaPro----- 212  
 Qy 409 GAACTATGAGCGGTACAGTCACCTCTGTGTTACTTGGCTGCGCTCAACCGTCAGGCTGC 468  
 Db 213 -----GlyProGlnGlyProProGly-----AlaProGlyGlnProGlySerGly 227  
 Qy 469 CACAGCTGAACCTCCGAGCTAGCTGCGCCACTTCTGTACAGGCTCAGGCGCTGCTGGG 528  
 Db 228 AsnAla-----ProGlyProAla-----GlyProProGlyProAlaGly 240  
 Qy 529 CAG-----CATTCAGGTGTATGGCGAGCTTGGCTTACCCACTGCCCA 573  
 Db 241 ProAsnGlyGlnProGlyHis-ProGlyGlnAspGlyGlnProGlyAlaProGlyAsnAs 260  
 Qy 574 GCCTCTGCCAGGACTGAGCAGCTGGGCGCCCTGGCCCTGCCACAGTGACTTCTCCCA 633  
 Db 260 pGlyThrProGlySerAspAlaAlaGlyCysProCysProThrArgSerValLeuAr 280  
 Qy 634 G 634  
 Db 280 g 280  
 RESULT 13  
 ID CC12\_CAEEL STANDARD; PRT; 316 AA.  
 AC P20630;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cuticle collagen 12 precursor.  
 GN COL-12 OR F15H10.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=90172409; PubMed=1689778;  
 RA Park Y.-S., Kramer J.M.;  
 RT "Tandemly duplicated Caenorhabditis elegans collagen genes differ in  
 RT their modes of splicing";  
 RL J. Mol. Biol. 211:395-406(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 CC LINKS.  
 CC -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 DR EMBL; X51622; CAA35954.1; -;  
 DR EMBL; Z73972; CAA98257.1; -;  
 DR PIR; S08169; S08169.  
 DR WormPep; F15H10.1; CE05638.  
 DR InterPro; IPR002486; Col\_cuticle\_N.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen; 3.  
 DR Pfam; PF01484; Col\_cuticle\_N; 1.  
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;  
 KW Signal.  
 FT SIGNAL 1 36 POTENTIAL.  
 FT CHAIN 37 316 CUTICLE COLLAGEN 12.  
 FT DOMAIN 128 157 TRIPLE-HELICAL REGION.  
 FT DOMAIN 176 202 TRIPLE-HELICAL REGION.  
 FT DOMAIN 206 235 TRIPLE-HELICAL REGION.  
 FT DOMAIN 240 265 TRIPLE-HELICAL REGION.  
 FT DOMAIN 269 304 TRIPLE-HELICAL REGION.  
 SQ SEQUENCE 316 AA; 30098 MW; 6CA81FF94706D42E CRC64;  
 Alignment Scores:  
 Pred. No.: 0.0786 Length: 316  
 Score: 127.50 Matches: 64  
 Percent Similarity: 33.19% Conservative: 12  
 Best Local Similarity: 27.95% Mismatches: 66  
 Query Match: 8.72% Indels: 87  
 DB: 1 Gaps: 13  
 US-09-931-704-4 (1-819) x CC12\_CAEEL (1-316)  
 Qy 638 TCCTCTGAGGAAGTCACTGTGGCGAGGCGCCAGGCCCGCCAGG---CTGGCTCAGTCCCTG 582  
 Db 117 SerSerGlyGlySerCysCysSerGlySerGlyAlaAlaGlyProAlaGlySerPro 136  
 Qy 581 GCAGAGGCTGGCGAGTGGGTAGCCAGCGTCGCCA----- 546  
 Db 137 GlyGlnAspGlyAlaProGlyAsnAspGlyAlaProGlyAlaProGlyAsnProGlyGln 156  
 Qy 545 -----TGACACCTGCATATGC-----TGCCAGCAGGCG 519  
 Db 157 AspAlaSerGluAspGlnThrAlaGlyProAspSerPheCysPheAspCysProAlaGly 176  
 Qy 518 CTTGGAGGCTGTACAGAGTGGGCGCAGCTACGTCGGAGTTCACTGTGGCAGCCTGAC 459  
 Db 177 Pro-----ProGlyProSerGlyAlaPro----- 184  
 Qy 458 GGTGAGGCCACGCAAGTAACACAGGAGGTGACTGTACGCTCATAGTTCTTGGGTGAGCC 399  
 Db 185 -----GlyGlnLysGlyProSerGlyAlaProGlyAlaProGlyGlnSerGlyGlyAla 202  
 Qy 398 GCAGCCTGTCTATTGAGGCTTCCCACTCCACATTCAAGTTGACCGTGGCCCTGGCAGAGTTT 339  
 Db 203 Ala-----LeuProGlyProProGlyPro----- 210  
 Qy 338 CTGCCCCCAGTCGAGGAGATTGAAGTCAGGCTCGTTGAAGAGGGGCGCCAGCTAGTTCA 279  
 Db 211 -----AlaGlyProProGlyProAla 217  
 Qy 278 GGTAGGTCCAGCTAAGTCGCGAGTTGATGCTCCAGGTAAGCGGTGAGGTCTAGGTTT 219  
 Db 218 GlyGlnProGlySerAsnGlyAsnAlaGlyAlaProGlyAla----- 231  
 Qy 218 TCTGGATGAGGGGCGCAGGCGCTCGATCTCTGTGCGATTAGAGCTGGCAGCTGCAGGGA 159  
 Db 232 -----ProGlyGlnValValAspVal-----Pro 239  
 Qy 158 GGTGCCACGACCGTGCATAGCAAGCTAACATCCCCACGAGTCCCTGCTCGGAGGT 99

Db 240 GlyThrProGlyProAla-----GlyProProGlySerPro-----Gly 252

QY 98 CCATGGGCTGGGGCGCGCGCGCGCGCGCTCTCTCCACAGAGCTGGCGGAGTG 39

Db 253 ProAlaGly---AlaProGlyGlnProGlyGlnAlaGlySerSerGlnProGlyGlyPro 271

QY 38 GGA-----GGCGAGCGCGGCTCCG 18

Db 272 GlyProGlnGlyAspAlaGlyAlaPro 280

RESULT 14

CC13\_CABEL STANDARD; PRT; 316 AA.

AC P20631;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Cuticle collagen 13 precursor.

GN COL-13 OR F15H10.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

ON NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=90172409; PubMed=1689778;

RA Park Y.-S., Kramer J.M.;

RT "tandemly duplicated Caenorhabditis elegans collagen genes differ in

RT their modes of splicing.";

RL J. Mol. Biol. 211:395-406(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Berks M.;

RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE

CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A

CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.

CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE

CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-

CC LINKS.

CC -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; X51623; CAA35955.1; -;

DR EMBL; Z73972; CAA98258.1; -;

DR PIR; S08170; S08170.

DR WormPep; F15H10.2; CE05639.

DR InterPro; IPR002486; Col\_cuticle\_N.

DR InterPro; IPR000087; Collagen.

DR Pfam; PF01391; Collagen; 3.

DR Pfam; PF01484; Col\_cuticle\_N; 1.

KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;

KW Signal.

FT SIGNAL 1 36

FT CHAIN 37 316

FT DOMAIN 128 157

FT DOMAIN 176 202

FT DOMAIN 206 235

FT DOMAIN 240 266

FT DOMAIN 269 304

FT DOMAIN 316 316

SQ SEQUENCE 316 AA; 30100 MW; 00C6D08FBC4701AF CRC64;

Alignment Scores:

Pred. No.: 0.0786 Length: 316

Score: 127.50 Matches: 64

Percent Similarity: 33.19% Conservative: 12

Best Local Similarity: 27.95% Mismatches: 66

Query Match: 8.72% Indels: 87

DB: 1 Gaps: 13

US-09-931-704-4 (1-819) x CC13\_CABEL (1-316)

QY 638 TCCTTGGGGAAGTCACTCTGGCGAGGCGCGAGGCGCCAGG---CTGGCTCAGTCCTGTG 582

Db 117 SerSerGlyGlySerCysSerCysGlySerGlyAlaAlaGlyProAlaGlySerPro 136

QY 581 GCAGAGGCTGGGCGAGTGGGTACCAAGCTCGCCA----- 546

Db 137 GlyGlnAspGlyAlaProGlyAsnAspGlyAlaProGlyAlaProGlyAsnProGlyGln 156

QY 545 -----TGACACCTGCAATGC-----TGCCCGCAGCAGGC 519

Db 157 AspAlaSerGluAspGlnThrAlaGlyProAspSerPheCysPheAspCysProAlaGly 176

QY 518 CCTGGAGGCTGGTACAGAAGTGGCGCAGGCTACGTGGAGTTGAGTGTGGCAGCCTGAC 459

Db 177 Pro-----ProGlyProSerGlyAlaPro----- 184

QY 458 GGTGTAGGCGCACGCAAGTAACACAGGAGGTGACTGTACGCTCATAGTTCTGGTGCAGCC 399

Db 185 -----GlyGlnGlySerGlyAlaProGlyAlaProGlyGlnSerGlyGlyAla 202

QY 398 GAGCCTGTCAATTGAGGCTTGGCCACACCTTCCAAGTTGACCGTGGCCCTGGGCGAGGTTT 339

Db 203 Ala-----LeuProGlyProProGlyPro----- 210

QY 338 CTGCCCCCAGTCGAGGAGGATTGAAGTCAGGCTCGTTGAAGGGGCGCCAGGAGTAGTTCA 279

Db 211 -----AlaGlyProGlyProAla 217

QY 278 GGTAGTCCAGCTAAGCTGCGGAGTTGATGCTCCAGGTAGCGGTTGAGGTGATAGTTT 219

Db 218 GlyGlnProGlySerAsnGlyAlaGlyAlaProGlyAla----- 231

QY 218 TCTGTATGAGGGGCGCGGCGGCTGATCTCTGTGCGATTAAAGAGCTGGCAGTGCAGGGA 159

Db 232 -----ProGlyGlnValValAspVal-----Pro 239

QY 158 GGTGCCACAGCACCGTGCATAGGCAAGCTAAACATCCCCACAGGTCCCTCTGCGGAGGT 99

Db 240 GlyThrProGlyProAla-----GlyProProGlySerPro-----Gly 252

QY 98 CCATGGGCTGGGCGCGCGCGCGCGGCTCTCTCCAGAGGCTGGCGGAGTG 39

Db 253 ProAlaGly---AlaProGlyGlnProGlyGlnAlaGlySerSerGlnProGlyGlyPro 271

QY 38 GGA-----GGCGAGCGCGGCTCCG 18

Db 272 GlyProGlnGlyAspAlaGlyAlaPro 280

RESULT 15

NME3 RAT

ID NME3 RAT STANDARD; PRT; 1237 AA.

AC Q00961;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Glutamate [NMDA] receptor subunit epsilon 3 precursor (N-methyl

GN GRIN2C.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:48:57 ; Search time 51.2562 Seconds  
(without alignments)  
6584.667 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 1498

Sequence: 1 tattattaaagtctccgg.....agccacgtcagctgtgctt 819

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09931704/runat\_27012003\_154126\_3635/app\_query.fasta\_1.7189  
-DB=SPTRMBL\_21 -QPMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pt -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pt -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09931704@cgn\_1.1.633 @runat\_27012003\_154126\_3635 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1224	81.7	225	11 Q9QZM3	Q9qzm3 mus musculus

C	2	1193	79.6	225	4	Q9UBD9	Q9ubd9 homo sapien
C	3	157	10.7	373	5	Q9BIT9	Q9bit9 latrodectus
C	4	154.5	10.6	399	5	Q9BIT8	Q9bit8 latrodectus
C	5	154	10.3	889	16	Q9F2N5	Q9f2n5 streptomyce
C	6	149	10.2	342	5	Q9BIU1	Q9biu1 gasteracant
C	7	149	10.2	1953	5	Q9BIT7	Q9bit7 nephila mad
C	8	145	9.7	325	5	Q9BIU6	Q9biu6 caenorhabdi
C	9	144.5	9.6	215	13	Q9PUJ2	Q9puj2 plethodon j
C	10	144.5	9.6	215	13	Q9PUJ1	Q9puj1 plethodon j
C	11	144.5	9.6	215	13	Q9PUJ0	Q9puj0 plethodon j
C	12	143	9.5	585	12	O41935	O41935 murid herpe
C	13	142.5	9.5	215	13	Q9PUI9	Q9pu9 plethodon j
C	14	142.5	9.5	608	12	Q9Q5K9	Q9q5k9 herpesvirus
C	15	141	9.4	383	12	O4397	O4397 epstein-bar
C	16	141	9.4	1691	11	Q9ESQ2	Q9esq2 mus musculu
C	17	140	9.3	676	6	Q9SJC9	Q9sjc9 sus scrofa
C	18	139	9.3	511	6	Q9SJD0	Q9sjd0 sus scrofa
C	19	139	9.3	566	6	Q9SJD1	Q9sjd1 sus scrofa
C	20	138	9.2	406	2	Q8VPM8	Q8vpm8 micrococcus
C	21	137.5	9.2	437	12	Q69146	Q69146 human herpe
C	22	137	9.1	319	5	O17038	O17038 caenorhabdi
C	23	137	9.1	574	3	O36027	O36027 schizosacch
C	24	137	9.1	773	16	Q9LIR7	Q9lir7 streptomyce
C	25	137	9.1	886	4	Q9NUB7	Q9nub7 homo sapien
C	26	136.5	9.1	1146	13	Q90584	Q90584 gallus gall
C	27	136	9.1	297	5	Q9NFL9	Q9nfl9 globodera p
C	28	136	9.3	444	5	Q9BIU6	Q9biu6 argiope tri
C	29	135.5	9.0	1616	4	O15054	O15054 homo sapien
C	30	135	9.2	371	5	O45114	O45114 caenorhabdi
C	31	135	9.2	871	5	O44358	O44358 nephila cla
C	32	135	9.2	1884	5	Q9NHW2	Q9nhw2 nephila mad
C	33	134.5	9.0	327	12	Q69145	Q69145 human herpe
C	34	134	8.9	302	5	O19079	O19079 caenorhabdi
C	35	133.5	8.9	506	3	Q960P4	Q96up4 aspergillus
C	36	133	8.9	304	5	Q9XUE9	Q9xue9 caenorhabdi
C	37	132.5	8.8	1745	4	Q9NZQ6	Q9nqz6 homo sapien
C	38	132	9.0	604	16	Q9L252	Q9l252 streptomyce
C	39	132	9.0	651	5	Q9BIU9	Q9biu9 argiope tri
C	40	132	8.8	652	5	Q9ST9	Q9sc9 drosophila
C	41	132	8.8	702	5	Q9VSH4	Q9vsh4 drosophila
C	42	132	8.8	1344	13	Q93419	Q93419 gallus gall
C	43	132	8.8	1433	11	Q07563	Q07563 mus musculu
C	44	131.5	8.8	303	5	Q23364	Q23364 caenorhabdi
C	45	131.5	9.0	563	5	Q9BIT5	Q9bit5 nephila mad

ALIGNMENTS

RESULT 1

Q9QZM3	PRELIMINARY;	PRT;	225 AA.
ID	Q9QZM3		
AC	Q9QZM3		
DT	01-MAY-2000 (TRENBLrel. 13, Created)		
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)		
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)		
DE	Neurotrophin-1/B-cell stimulating factor-3.		
GN	BSF3.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99432254; PubMed=10500198;		
RA	Senaldi G, Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S.,		
RA	Guo J., Elliott G., McNinch J., Shuklee C.L., Freeman D., Manu F.,		
RA	Simonet W.S., Boone T., Chang M.-S.		
RT	"Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the		
RT	IL-6 family."		
RL	Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).		
DR	EMBL; AF176913; AAF00993.1; -		
DR	MGI; MGI:1930088; Bsf3.		
DR	SEQUENCE 225 AA; 25261 MW; 68B1FEAAB7F1A950 CRC64;		

## Alignment Scores:

Pred. No.: 1,2e-96 Length: 225  
 Score: 1224.00 Matches: 225  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 81.71% Indels: 0  
 DB: 11 Gaps: 0

US-09-931-704-4 (1-819) x Q9QZM3 (1-225)

QY 95 ATGACCTCCGAGCGGACGTCGTGGGGATGTTAGCTTGCTATGACGGTGTGTGG 154  
 DB 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 155 CACCTCCCTGAGTGCAGCTCTTAATCGCAGAGATCCAGGCCCTGCCCTCCATC 214  
 DB 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
 QY 215 CAGAAACCTATGACCTACCCGCTACCTGGAGCATCACTCCGACGCTTAGCTGGACC 274  
 DB 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 275 TACCTGAACCTACCTGGGCCCCCTTTCAACGAGCCCTGACTTCAATCCTCTCGACTGGG 334  
 DB 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 QY 335 CGAGAACTCTGCCAGGCCGACGCTCACTTGGAGTGTGGGAGACCTCAATGACAGG 394  
 DB 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100  
 QY 395 CTGCGCTGACCCAGCAATGAGGCGTACAGTCACCTCCTCTGTACTTGGCTGGGCTC 454  
 DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 455 AACCGTCAAGCTGACCTGGAACCTCCGAGTAGCTGGCCCACTTCTGTACAGCCTC 514  
 DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 515 CAGGCGCTGCTGGGAGCAATGAGGTGTCATGGCAGCGCTGGCTACCCAGCCCCAG 574  
 DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160  
 QY 575 CCTCTCCAGGAGTCCAGCAGCGCTGGGCCCTGGCCCTCCACAGTACTTCTCTCCAG 634  
 DB 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180  
 QY 635 AAGATGATGACTTCTGGTGTCTAAGGAGCTGCAGACCTGGCTATGGCGTTCAGCCAAG 694  
 DB 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 695 GACTTCAACGGCTTAAGAGAGATGACGCTCCAGCAGCTTCAGTCACCTGCACCTTG 754  
 DB 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
 QY 755 GAGGCACATGGTTTC 769  
 DB 221 GluAlaHisGlyPhe 225

## RESULT 2

Q9UBD9 PRELIMINARY; PRT; 225 AA.  
 ID Q9UBD9; PRT; 225 AA.  
 AC Q9UBD9;  
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
 DE Neurotrophin-1/3-cell stimulating factor-3 (Cardiotrophin-like  
 DE cytokine) (Similar to cardiotrophin-like cytokine,  
 DE neurotrophin-1/3-cell stimulating factor-3).  
 GN C1C.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI\_TaxID=9606;

OX [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99432254; PubMed=10500198;  
 RA Senaldi G., Varnum B.C., Samiento U., Starnes C., Lile J., Scully S.,  
 RA Guo J., Elliott G., McIninch J., Shaklee C.L., Freeman D., Manu F.,  
 RA Simonet W.S., Boone T., Chang M.-S.;  
 RT "Novel neurotrophin-1/3 cell-stimulating factor-3: A cytokine of the  
 RT IL-6 family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99382254; PubMed=10448081;  
 RA Shi Y., Wang W., Yourey P.A., Gohari S., Zukauskas D., Zhang J.,  
 RA Ruben S., Alderson R.P.;  
 RT "Computational EST database analysis identifies a novel member of the  
 RT neurotrophic cytokine family.";  
 RL Biochem. Biophys. Res. Commun. 262:132-138 (1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX Hu X., Xu Y., Zhang B., Peng X., Yuan J., Qiang B.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=KIDNEY;  
 RA Strausberg R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF176912; AAF0092.1; -;  
 DR EMBL; AF172854; AAD54284.1; -;  
 DR EMBL; AF176911; AAF00991.1; -;  
 DR EMBL; AY049779; AAL15436.1; -;  
 DR EMBL; BC012939; AAH12939.1; -;  
 SQ SEQUENCE 225 AA; 25176 MW; E2DD4B6280833B55 CRC64;

## Alignment Scores:

Pred. No.: 5,45e-94 Length: 225  
 Score: 1193.00 Matches: 218  
 Percent Similarity: 98.22% Conservative: 3  
 Best Local Similarity: 96.89% Mismatches: 4  
 Query Match: 79.64% Indels: 0  
 DB: 4 Gaps: 0

US-09-931-704-4 (1-819) x Q9UBD9 (1-225)

QY 95 ATGACCTCCGAGCGGACGTCGTGGGGATGTTAGCTTGCTATGACGGTGTGTGG 154  
 DB 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 155 CACCTCCCTGAGTGCAGCTCTTAATCGCAGAGATCCAGGCCCTGCCCTCCATC 214  
 DB 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
 QY 215 CAGAAACCTATGACCTACCCGCTACCTGGAGCATCACTCCGACGCTTAGCTGGACC 274  
 DB 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 275 TACCTGAACCTACCTGGGCCCCCTTTCAACGAGCTGACTTCAATCCTCTCGACTGGG 334  
 DB 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 QY 335 CGAGAACTCTGCCAGGCCGACGCTCACTTGGAGCATCACTCCGACGCTTAGCTGGACC 394  
 DB 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
 QY 395 CTGCGCTGACCCAGCAATGAGGCGTACAGTACCTCCTGTGTACTTGGCTGGCCTC 454  
 DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 455 AACCGTCAAGCTGACCTGGAACCTCCGAGTAGCTGGCCCACTTCTGTACAGCCTC 514  
 DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 515 CAGGCGCTGCTGGGAGCAATGAGGTGTCATGGCAGCGCTGGCTACCCAGCCCCAG 574













```

Db 75 LeuProHisIleLysValAlaAsnLeuProThrAlaAlaMetAspTyrAspThrPheMet 94
Qy 380 AGCTCAATGACAGCTGGGTGACCCAGAACTATGAGGCGTACAGTACCTCCTGTGT 439
Db 95 LysGlnThrAspGluThrArgLeuAsnAsnLeuTyrPheTyrSerAlaIleValGlu 114
Qy 440 TACTTGGCT---GGCTCAACGCTCAG-----GCTGCCACAGCTGAACCTCGA 484
Db 115 PheLeuLysGluAlaMetThrGluGlnGluAspLeuAsnProAlaGluLeuSerLeuLys 134
Qy 485 CQTGCTGGCCCACTCTGTACACGCTCCAGGCTCCAGGCGCTGCTGGCGACATTCGAGGTGC 544
Db 135 AlaLysPheGluGluAlaMetAlaAsnSerAsnThrLeuIleSerLysIleSerAspIle 154
Qy 545 ATGGCGACGCTGGTACCCACCTGCCCCAGCTCTGCCAGGAGTACGAGCAGCTGGGCC 604
Db 155 MetThrGlnMetGlyMetSerValThrIleThrLeuPro-----LysProLeuValVal 172
Qy 605 CTTGCCCTGCCACAGTACTCTCTCCAGAGATGGATGACTTCTGGCTGTGAAGGAG 664
Db 173 ProPheGluGlySerAlaTyrPheArgLysLysLeuArgGlyGlyValValCysLysGlu 192
Qy 665 CTGCAGACCTGGCTATGGGCTTCAGCCAGAGACTTCAACCGGCTTAAGAAGAAGATCCAG 724
Db 193 TyrLysGluArgValLeuLeuThrLysArgAspPheGluPheLeuAlaLysLysTyrGln 212

RESULT 11
Q9PUJO
ID Q9PUJO PRELIMINARY; PRT; 215 AA.
AC Q9PUJO;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Receptivity factor isoform 3 precursor.
GN PRF.
OS Plethodon jordani (Salamander).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Plethodontidae;
OC Plethodon.
OX NCBI_TaxID=8336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420364; PubMed=10489368;
RA Pollmann S.M., Houck L.D., Feldhoff R.C.;
RT "proteinaceous pheromone affecting female receptivity in a terrestrial salamander.";
RL Science 285:1907-1909 (1999).
DR EMBL; AF181482; AAF01027.1; -.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 1 23 POTENTIAL.
Alignment Scores:
Pred. No.: 0.000428 Length: 215
Score: 144.50 Matches: 44
Percent Similarity: 46.99% Conservative: 34
Best Local Similarity: 26.51% Mismatches: 73
Query Match: 9.65% Indels: 15
DB: 13 Gaps: 4

US-09-931-704-4 (1-819) x Q9PUJO (1-215)
Qy 260 AGCTTAGTGGGACCTACTCACTACCTGCGGCGCCCTTTCAACGAGCGCTGACTTCAAT 319
Db 55 SerLeuLeuProThrTyrLeuSerPheGlnGlyAlaProLeuSerAspProAspTyrGln 74
Qy 320 CCTCTCGACTGGGGCAGAACTCTGCCAGGGCCACGGTCAACTGGAAGTGTGGCA 379
Db ,75 LeuProHisIleLysValAlaAsnLeuProThrAlaAlaMetAspTyrAspThrPheIle 94
Qy 380 AGCTCAATGACAGCTGGGTGACCCAGAACTATGAGGCGTACAGTACCTCCTGTGT 439
Db 95 ArgGlnThrAspGluThrArgLeuLysAsnLeuTyrPheTyrSerAlaIleValGlu 114

```

```

Qy 440 TACTTGGCT-----GGCTCAAC-----CGTCAGGCTGCCACA 472
Db 115 PheLeuLysGluAlaMetThrGluGlnGluAspLeuAsnProAlaGluLeuAlaLeuLys 134
Qy 473 GCTGAACCTCGAGTACGCTGGCCACCTCTGTATACCAAGCTCCAGGGCGCTGCTGGCAGC 532
Db 135 AlaLysPheGluGluAlaMetAlaAsnSerAsnThr-----LeuIleSerLys 150
Qy 533 ATTGCAGGTCTCATGGCGACGCTGGTGTACCTACCTGCCCCAGCTCTGCCAGGGGACTGAG 592
Db 151 IleSerAspIleMetThrGlnMetGlyMetSerValThrIleThrLeuPro-----Lys 168
Qy 593 CCAGCTCGGGCCCTGGCCCTCCACAGTACGATTCCTCCAGAAATGATGATCTCTTGG 652
Db 169 ProLeuValValProPheLysGlySerAlaTyrPheSerLysLysLeuArgGlyGlyVal 188
Qy 653 CTGCTGAAGAGCTGCAGACCTGGCTGATGGCTTGCAGCAAGGACTTCAACCGGCTTAAG 712
Db 189 ValCysLysGluTyrLysGluArgValPheLeuThrLysArgAspPheMetLeuLeuAla 208
Qy 713 AAGAAGATGCAGCTCCA 730
Db 209 GluLysTyrGlnGlyPro 214

RESULT 12
O41935
ID O41935 PRELIMINARY; PRT; 585 AA.
AC O41935;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE Hypothetical 60.2 kDa protein.
GN GAMAHV.M6.
OS murid herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=33708;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=WUMS;
RX MEDLINE=97366649; PubMed=9223479;
RA Virgin H.W. IV, Latreille P., Wameley P., Halleworth K., Weck K.E.,
RA Dal Canto A.J., Speck S.H.;
RT "Complete sequence and genomic analysis of murine gammaherpesvirus 68.";
RL J. Virol. 71:5894-5904 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=WUMS;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U97553; AAB66392.1; -.
KW Hypothetical protein.
SQ SEQUENCE 585 AA; 60160 MW; 85610AA80C34827D CRC64;

Alignment Scores:
Pred. No.: 0.00067 Length: 585
Score: 143.00 Matches: 75
Percent Similarity: 35.74% Conservative: 14
Best Local Similarity: 30.12% Mismatches: 82
Query Match: 9.55% Indels: 78
DB: 12 Gaps: 12

US-09-931-704-4 (1-819) x O41935 (1-585)
Qy 19 GGAGCCGGCTCGCCCTCCACTCCG-----CCAGCTCTGGGAGAGGA 63
Db 334 GlyProAspProArgProProArgProProGluLeuGlyProGlySerProThrSer 353
Qy 64 GCGCCGCCCGCGCGG----- 78
Db 354 ProAlaProSerArgAlaGlyAlaArgIleProAspLeuProGlyProLeuProSerTrp 373

```

```

QY 79 ---CCCGGCCCCAGCCCATGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTTG 135
Db 374 GlyProAspProArgProProProProProProProProProProProProProPro 383
QY 136 CCTATGCACGGTGTGGGACCTCCCTGCGAGTCCAGCTCTTAATCCACAGAGAGA--- 192
Db 384 ProGluLeuGlyProGlySerProThrSerProAlaProSerArgAlaGlyAlaArgIle 403
QY 193 ---TCCAGGCCCTGGCCCTC-----CATCCAGAAAACCTATGACCTCACCCCGCTA 240
Db 404 ProAspLeuProGlyProLeuProSerTrpGlyProAspProArgProProArgProPro 423
QY 241 CCT-----GGAGCATCAACT 255
Db 424 ProGluLeuGlyProGlySerProThrSerProAlaProSerArgAlaGlyAlaArgIle 443
QY 256 CCGCAGCTTAGCTGGGACCTACCTGAACCTA---CCTGGGCCCC---TTTCAACGAGCC 309
Db 444 ProArgProSerArgProProProGluLeuGlyProGlySerProArgPheProArg-Pr 463
QY 310 TGACTTCAATCTCTCCGACTG-----GGGGCAGAAACTCTGCCAGGGCCACGGTCAA 363
Db 463 oSerArgProProGluLeuGlyProGlySerProAspLeu-ProGlyProLeuProS 483
QY 364 CTTGGAGTGTGGGAGCTCTCAATGACAGCTCGGCTGACCCAGAACTATGAGGCGTA 423
Db 483 erTrp-GlyProAspProProThrPheProAla-----ProSerArgAlaGlyAla 499
QY 424 CAGTCACCTCTGTGTACTTGGTGGCTCAACCGTCAAGCTGCGGCTGCCACAGTCACTCCG 483
Db 500 ArgIleProProValProProThrPheProAlaProSerArgAlaGlyAlaArgIlePro 519
QY 484 ACTAGCTGGCCCACTTCTGTACCACTCCAGGGCTGCTGGGAGCAATTCAGGTGT 543
Db 520 ArgProPro-----ArgProProGluLeuGly----- 529
QY 544 CATGGCGAGCTTGGTACCCAGTGGCCCGGCTCTGCGAGGAGCTGAGCCAGCTGGGC 603
Db 530 -----ProGlySerPro-AspLeuProGlyProLeuProSerTrpG1 543
QY 604 CCTGGCCCTGCCACAGT 622
Db 543 yProValProProProSer 549

RESULT 13
Q9PUI9
ID Q9PUI9 PRELIMINARY; PRT; 215 AA.
AC Q9PUI9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Receptivity factor isoform 4 precursor.
GN PRF.
OS Plethodon jordani (Salamander).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidae; Plethodontidae;
OC Plethodon.
OX NCBI_TaxID=8336;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99420364; PubMed=10489368;
RT "Proteinaceous pheromone affecting female receptivity in a terrestrial
RT salamander.";
RL Science 285:1907-1909 (1999).
DR EMBL; AF181483; AAF01028.1; -.
KW SIGNAL.
FT SIGNAL
SQ SEQUENCE 215 AA; 24054 MW; AA412135FFA4E7F CRC64;

Alignment Scores:
Pred. No.: 1 23 POTENTIAL.
Score: 142.50
Percent Similarity: 31.07%
Best Local Similarity: 27.12%
Query Match: 9.51%
Indels: 113
Gaps: 12
Length: 215

```

```

Score: 142.50 Matches: 41
Percent Similarity: 45.68% Conservative: 33
Best Local Similarity: 25.31% Mismatches: 81
Query Match: 9.51% Indels: 7
DB: 13 Gaps: 3

US-09-931-704-4 (1-819) x Q9PUI9 (1-215)

QY 260 AGCTTACTGGGACCTACCTAGTAACCTGCGGCCCCCTTTCAACGAGCCTGACTTCAAT 319
Db 55 SerLeuLeuProThrThrLeuSerPheGlnGlyAlaProLeuSerAspProAspTyrArg 74
QY 320 CCTCTCACTGCTGGGGCAGAAACTCTCCAGGSCCAGCGTCAACTTTGGAAGTGGCGGA 379
Db 75 LeuProHisIleLeuValValAsnLeuProThrAlaAlaMetAspTyrAspThrPheIle 94
QY 380 AGCTTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGCGTACAGTCACTCTGTGT 439
Db 95 ArgGlnThrAspGluThrArgLeuLeuAsnAsnLeuTyrPheTyrSerAlaIleValGlu 114
QY 440 TACTTGGT---GGCCTCAACCTGAG-----GCTGCCACAGCTGAACCTCGA 484
Db 115 PheLeuLeuGluAlaMetThrGluGlnGluAspLeuAsnProAlaGluLeuSerLeuLys 134
QY 485 CGTAGCTGCGCCACTTCTGTACGAGCTCCAGGGCTGCTGGGCGAGCATTTGCAGGTC 544
Db 135 AlaLysPheGluGluAlaMetAlaAsnSerAsnThrLeuLeuSerLysIleSerAspIle 154
QY 545 ATGGCAGCGTGGTACCCACTGCCCCAGCCTCTGCCAGGAGTACAGCCAGCTGGGCC 604
Db 155 MetThrGlnMetGlyMetSerValThrIleThrLeuPro-----LysProLeuValVal 172
QY 605 CTGGGCGCTGCCCCACAGTACTCTCCAGAACTGATGATGATCTCTGGCTGCTGAAGAG 664
Db 173 ProPheLysGlySerAlaTyrPheSerLysLeuValGlyValValCysLeuGlu 192
QY 665 CTCAGACCTGGCTATGGCTTCAGCAAGGACTTCAACCGGCTTAAGAAGAGAGATCAG 724
Db 193 TyrLysGluArgValPheLeuThrLysArgAspPheMetLeuLeuAlaGluLysTyrGln 212
QY 725 CTCTCA 730
Db 213 GlyPro 214

RESULT 14
Q9Q5K9
ID Q9Q5K9 PRELIMINARY; PRT; 608 AA.
AC Q9Q5K9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE NTR.
OS Herpesvirus papio.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10394;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BABOON LYMPHOCRYPTOVIRUS BA65;
RA Zong J.-C., Ryan J., Ling F.D., Loeb D.D., Pagano J.S., Hayward S.D.,
RA Hayward G.S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF200364; AAF23950.1; -.
SQ SEQUENCE 608 AA; 60458 MW; 4BF82ACB0A029627 CRC64;

Alignment Scores:
Pred. No.: 0.000744 Length: 608
Score: 142.50 Matches: 96
Percent Similarity: 31.07% Conservative: 14
Best Local Similarity: 27.12% Mismatches: 113
Query Match: 9.51% Indels: 131
DB: 12 Gaps: 18

```

```
US-09-931-704-4 (1-819) x Q905K9 (1-608)
QY 19 GGAGCGCGGCTCGCCTCCCACTCGCAGCCTCTGGAGAGAGCGCGCGCGCGG 78
Db 68 GlyAlaGlyGlnArg--ProSerGlyProThrGlyGlyHisProAlaAlaProGlyAla 86
QY 79 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 99
Db 87 ProGlyProArgSerProArgThrGluArgArgGlySerAlaGlnArgGlyHisPro 106
QY 100 CCGCGAGCAGCGGACTCGTGGGGATGTAGCTTCCTATGACGCTGCTGGCAGCT 159
Db 107 ProProGlyAlaGly-----GlnArgProSerGlyProThrGlyGlyHisProAlaAla 124
QY 160 CCTCGAGTGCAGCTCTTAATCGCAGAGATCAGGCGCTGGCGCTCCATCCAGAA 219
Db 125 ProGlyAlaProGlyProArgSerProArgThrGluArgArgGlySerAlaGlnArg 144
QY 220 AACCTATGACCTACCGCTACCTGGAGCA-----TCAACTCCGCGAGCTTAGC 267
Db 145 Gly-----HisProProGlyAlaGlyGlnArgProSerGlyPro-----Thr 159
QY 268 TGGGACCTACCTGAACTACCTGGGGCGCGCGCGCGCGCGCGCGCGCGCG 327
Db 160 GlyGlyHisProAlaAlaProGlyAlaProGlyProArgSer-----ProArgThr 176
QY 328 ACTGGGGGAGAACTCTGCCAG-----GGCAGCGGTCAACTTGGAGTGGCGAGCCT 384
Db 177 GluArgArgGlySerAlaGlnArgGlyHisProProProGlyAlaGlyGlnArgPro 196
QY 385 CAATGACAGCTCGCGCTGACCCAGAACTATGAGCGCTACAGTCACCT----- 432
Db 197 SerGlyProThrGlyGlyHisProAlaAlaProGlyAlaProGlyProArgSerProArg 216
QY 433 -----CCTGTGTACTTGGTGGCGCTCA 456
Db 217 ThrGluArgArgGlySerAlaGlnArgGlyHisProProProGlyAlaGlyGlnArg 236
QY 457 CCGTCA-----GCCTGCCAGCTGAACCTCGAGCTAGCTGGCGCCACTTCTG--- 504
Db 237 ProSerGlyProThrGlyGlyHisProAlaAlaProGlyAlaProGlyProArgSerPro 256
QY 505 -----TAC 507
Db 257 ArgThrGluArgArgGlySerAlaGlnArgGlyHisProProProGlyAlaGlyGln 276
QY 508 CAGCCTCCAGCGCTGTGGCAGCATTCAGGTGTATGCGGAGCTTGGCTACCCACT 567
Db 277 ArgProSerGlyProThrGlyGlyHis----- 285
QY 568 GCCCGACCTCTGCAGGACTGAGCCAGCTGGGCGCGCTGGCGCGCTGGCGCGCTG 627
Db 286 ---ProAlaAla-ProGly-----AlaProGlyProArg-SerProArgT 299
QY 628 CCTCCAGAGATGATGATCTTCTGCTGCTGAAGAGCTGCAGA----- 671
Db 299 hrGluArgArg-----ArgGlySerAlaGlnArgGlyHisProp 312
QY 672 --CTGGCTATGGCTTCAGCAGGACTTCAACCGCG----- 707
Db 312 roProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyHisProAlaAlaProGlyA 332
QY 708 -----TTAAGAAGAGATGACGCTCCAG----- 731
Db 332 laProGlyProArgSerProArgThrGluArgArgGlySerAlaGlnArgGlyHisP 352
QY 732 -----CAGCTTCACTCACCTCGCTGAGGAGGACATGGTTCTGACC 774
Db 352 roProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyHis----- 367
QY 775 TCTGACCTTAACCCACACCTCCAGGCCAGCTAGC 812
Db 368 ---ProAlaAlaProGlyAlaProGlyProArgSer 378
```

```
RESULT 15
Q04397 PRELIMINARY; PRT; 383 AA.
AC Q04397; 1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE Hypothetical BWRFL protein.
GN BWRFL OR BCRF2
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211 (1984).
[2]
SEQUENCE OF 88-145 FROM N.A.
RP STRAIN=B95-8;
RX MEDLINE=88296424; PubMed=2841116;
RA Walls D., Gannon F.;
RT "The expression of novel antigens from the Epstein-Barr virus large
internal repeat";
RL EMBO J. 7:1191-1196 (1988).
DR EMBL; V01555; CAA24870.1; -
DR EMBL; V01555; CAA24871.1; -
DR EMBL; V01555; CAA24872.1; -
DR EMBL; V01555; CAA24873.1; -
DR EMBL; V01555; CAA24874.1; -
DR EMBL; V01555; CAA24869.1; -
DR EMBL; V01555; CAA24868.1; -
DR EMBL; X07816; CAA30675.1; -
DR EMBL; V01555; CAA24864.1; -
DR EMBL; V01555; CAA24865.1; -
DR EMBL; V01555; CAA24866.1; -
DR EMBL; V01555; CAA24867.1; -
DR EMBL; V01555; CAA24868.1; -
KW Hypothetical protein.
SQ SEQUENCE 383 AA; 39866 MW; 3FA433CE1EB04461 CRC64;
Alignment Scores:
Pred. No.: 0.000933 Length: 383
Score: 141.00 Matches: 92
Percent Similarity: 31.96% Conservatives: 17
Best Local Similarity: 26.98% Mismatches: 97
Query Match: 9.41% Indels: 135
DB: 12 Gaps: 22
US-09-931-704-4 (1-819) x Q04397 (1-383)
QY 16 GCCGGAGCGCGCTCGCCTCCCACTCCGCGCA----- 48
Db 55 SerGlyAlaAlaGlySerProGlnAlaProGlnThrArgValSerAlaSerArgAla 74
QY 49 -----GCCTCTGGGAGAGGA----- 63
Db 75 AspArgProArgAlaTrpArgLeuLeuGlyAlaSerArgGlyTrpPheCysProSer 94
QY 64 -----GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 108
Db 95 LeuCysProSerGluGluProGlyThrSerGlyThrProGluProLeuGlyProAlaSer 114
QY 109 AGG-----GGACTCTGGGGGATTTAGCTTGGCTATGCAC----- 144
Db 115 ArgArgProGlyLeuArgSerProLeuSer---ProVallyProLysGluCysLeu 133
QY 145 -----GGTGTCTGGCGACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 189
```



```
Db 134 ArgGlyAlaThrLeuGlyAlaGlnAlaProGluSerArgGlyGln-----GlyHisLeu 151
QY 190 AGATCCAGGCC-----|||||
Db 152 ArgValProProArgValProGlyGlnProGluGlyProArgGlnProGlyArgProGln 171
QY 220 AACCTATGACCTACCGCTACCTGGAGCATCAACTCGCAGCTTAGCTGGGACCTACCT 279
Db 172 ArgProValProArgProGlyLeuGlnSerProGlyCysPro-----Pro 188
QY 280 GAA-----CTACCTGGGGCCCTTTCAACGAGCCTGACTTCAATCCTCCTCG 327
Db 189 GluGlyThrLeuGlyValProSerProPro-----LeuGlnAlaArgAla 203
QY 328 ACTGGGGGCAAACTCTGCCAGGGCCACGGTCAACTTGAAGTGTGGCGAAGCCT--- 384
Db 204 SerProSerArgGlyAla-----SerLeuGlyProGlnValGlnProHis 219
QY 385 CAATGACAGGCTGGGCTGACCCAGAACTATGAGGCGTACAGTCACTCTGTGTACTT 444
Db 220 ArgAspProSerGlyProAspProProThrGlyProSerLeuCysProProAlaProLeu 239
QY 445 GCCTGGCCTCAACCGTCAGGCTGCCACAGCTGAACCTCCGACGTAGCTGGCCACTTCTG 504
Db 240 -----GlnProSer---LeuHis-----ProArgProGlnLeu 249
QY 505 TACACGCTCCAGGCTGTGGGCGACATTGCAGGTGTATGGCGACGCTGGCTACCC 564
Db 250 LeuAlaSerProGlyProProGlyGln-----Pro 259
QY 565 ACTGCCCCAGCCTCTGCCAGGGACTGAGCCAGCTGGGCCCTGG-----609
Db 260 GluGlyProArgGlnProGlyArgValAlaPheProLeuProTrpProLeuProAla 279
QY 610 -----CCCTGCCACAGTGACTTCTCTCCA-----GAA 636
Db 280 SerHisProSerProLeuSerLeuProProHisArgValHisGlnAlaGlyArgArgAsp 299
QY 637 GATGGATGACTTCTGGCTGTGAAGAGCTGCAGACCTGGCTATGGCGTTACGCCAAGGA 696
Db 300 ProGly-----GlyProValSerValProProAlaAlaGlnSer 313
QY 697 CTTCAACCGCTTAAGAAGATGACGCTCCAGCAGCTTCACTCAGTCACTGCCTTGA 756
Db 314 LeuProPro-----GlyLysGlyAlaSerPheSer-----323
QY 757 GGCACATGGTTTCTGACCTTACCCCTTAACCCCAACCTCCAGGCCAGTCACTGCTG 816
Db 324 -----ProProSerLeu-ArgProSerLeuLeuCy 333
QY 817 C 817
Db 333 S 333
```

Search completed: January 27, 2003, 16:20:19  
Job time : 57.2562 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:43:12 ; Search time 34.4559 Seconds  
(without alignments)  
6334.602 Million cell updates/sec

Title: US-09-931-704-4  
Perfect score: 1498  
Sequence: 1 tattataaagcttcgcgg.....agggccagtcagctgtgctt 819

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09931704/runat\_27012003\_154125\_3585/app\_query.fasta\_1.7189  
-DB=A Geneseq 101002 -QMT=fastan -SUFFIX=tag -MINMATCH=0.1 -LOPCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLCN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09931704 @CEN\_1\_1\_222 @runat\_27012003\_154125\_3585 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 101002.\*

1:	/SIDS2/gcgdata/geneseq/genesep-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/genesep-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/genesep-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/genesep-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/genesep-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/genesep-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/genesep-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/genesep-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/genesep-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/genesep-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/genesep-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/genesep-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/genesep-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/genesep-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/genesep-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/genesep-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/genesep-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/genesep-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/genesep-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/genesep-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/genesep-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/genesep-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/genesep-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1324	88.4	253	22	AAW25831
2	1224	81.7	225	19	AAW29716
3	1224	81.7	225	19	AAW56142
4	1224	81.7	225	23	AAW7814
5	1224	81.7	225	23	AAW78177
6	1193	79.6	225	19	AAW29715
7	1193	79.6	225	19	AAW56141
8	1193	79.6	225	20	AAW94466
9	1193	79.6	225	21	AAW87813
10	1193	79.6	225	22	AAW63543
11	1193	79.6	225	23	AAW78176
12	1192	79.6	321	22	ABW11896
13	1192	79.6	321	22	AAW79399
14	1185	79.1	260	22	AAW78415
15	1181	78.8	223	22	AAW80828
16	1167	77.9	215	21	AAW19587
17	1136	75.8	164	21	AAW19586
18	852	56.9	164	22	ABB40317
19	852	56.9	164	22	ABB24716
20	852	56.9	164	22	AAW61118
21	852	56.9	164	22	AAW73827
22	852	56.9	164	22	AAW20115
23	852	56.9	164	22	AAW34012
24	852	56.9	164	23	ABG43716
25	161	10.7	208	20	AAW09196
26	159	10.6	208	20	AAW09197
27	146	9.7	330	19	AAW57645
28	146	9.7	408	17	AAW07539
29	146	9.7	408	22	AAW27377
30	146	9.7	408	22	AAW64007
31	145.5	9.7	689	20	AAW08306
32	142	9.5	325	22	ABG21919
33	141.5	9.4	371	22	ABG05200
34	140.5	9.4	1064	17	AAW93254
35	140.5	9.4	1064	19	AAW37741
36	140.5	9.4	1065	14	AAW37741
37	139	9.3	684	18	AAW15287
38	137	9.1	574	22	AAW52322
39	137	9.1	574	22	AAW67370
40	137	9.1	772	13	AAW23873
41	137	9.1	772	18	AAW09643
42	137	9.1	1685	22	ABG04839
43	137	9.1	1693	22	ABG15619
44	136	9.1	682	16	AAW80330
45	136	9.1	682	19	AAW49717

## ALIGNMENTS

RESULT 1

AAW25831

ID AAW25831 standard; Protein; 253 AA.

XX

AC AAW25831;

XX

DT 16-OCT-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:1346.

XX

Human; cancer; ulcer; HIV infection; human immunodeficiency virus; anti-infectious; antihemorrhagic; antithrombotic; antitumor; antiviral; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia; antiaggregant; haemostatic; vulnary; antiulcer; osteopathic; eczema; dermatological; antiallergic; antiaschmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antitense therapy; vaccine; inflammation; antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;

KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder.

XX Homo sapiens.

OS WO200153455-A2.

PN 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

PF 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

PI WPI: 2001-457603/49.

DR N-PSDB; AAH99772.

XX Isolated human polynucleotides encoding polypeptides, useful for the  
 treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 Claim 20; Page 278; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 CC AAM25963. The proteins can have activities based on the tissues and  
 CC cells they are expressed in, such as: antinflammatory; antirheumatic;  
 CC antithratic; immunosuppressive; antibacterial; endocrine; cardiac;  
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;  
 CC antitumor; osteopathic; dermatological; antiallergic; antisthmatic;  
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,  
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders.

XX SQ Sequence 253 AA;

Alignment Scores:  
 Pred. No.: 9.8e-114 Length: 253  
 Score: 1324.00 Matches: 245  
 Percent Similarity: 97.64% Conservative: 3  
 Best Local Similarity: 96.46% Mismatches: 5  
 Query Match: 88.38% Indels: 1  
 DB: 22 Gaps: 0

US-09-931-704-4 (1-819) x AAM25831 (1-253)

Qy 10 AGCTTGGCGGAGCGCGGCTCGCTCCACTCGCCAGCTCTGGGAGGAGCGCG 69  
 Db 1 SerPheAlaGlyAlaAlaAlaArgProSerThProAlaSerGlyArgGlyAlaAla 20  
 Qy 70 CCCG 129  
 Db 21 ProGlyArgProGly-ProSerProMetAspLeuArgAlaGlyAspSerTrpGlyMetLe 40

QY 130 AGCTTGGCTATGACCGGTGTGTGGCCCTCCCTGCTGAGTGCAGCTCTTAATCGCACAGG 189  
 Db 40 uAlaCysLeuCysThrValLeuTPhHisLeuProAlaValProAlaLeuAsnArgThrGl 60  
 QY 190 AGATTCAGGCGCTGGCGCCCTCCATCCGAAACCTATGACCTCACCCGCTACCTGGAGCA 249  
 Db 60 YAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGluHi 80  
 QY 250 TCAACTCCGACAGCTTAGCTGGGACCTACCTGAATACCTTGGGGCCCTTTTCAACGAGCC 309  
 Db 80 sGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPheAsnGluPr 100  
 QY 310 TGACTTCAATCCTCCTCGACTGGGGGAGAACTCTGCCAGGCGCACCGTCAACTTGA 369  
 Db 100 oAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeuGl 120  
 QY 370 AGTGTGGGAGAGCTCAATGACAGGCTCGGCTGACCCAGAGAACTATGAGCGGTACAGTCA 429  
 Db 120 uValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSerHi 140  
 QY 430 CCTCTGTGTACTTGTGGTGGCTCAACCGTCCAGCTGCCAGCTGCCAGCTGAACTCCGACGTAG 489  
 Db 140 sLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgArgSe 160  
 QY 490 CCTGGCCCACTTCTGACAGCTCCAGGCTCGGCTGCTGGGCGCATTTGCAGTGTCTATGCC 549  
 Db 160 rLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMetAl 180  
 QY 550 GACGCTGGCTACCCACCTGCGCCAGCTCTGCGCAGGAGCTGAGCCAGCTGGGCGCCCTGG 609  
 Db 180 aAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrProGl 200  
 QY 610 CCCTGCCCACTGATGCTTCTCCAGAGATGATGATGCTTCTGGCTCTGAGGAGCTGCA 669  
 Db 200 yProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeuGl 220  
 QY 670 GACCTGGCTATGGCTTCAGCCCAAGAGCTTCAACCGGCTTAAGAAGAAGATGAGAGCTCC 729  
 Db 220 nThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLeuMetGlnProPr 240  
 QY 730 AGCAGCTTCAGTCACCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 769  
 Db 240 oAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 253  
 RESULT 2  
 AAW29716  
 ID AAW29716 standard; Protein; 225 AA.  
 XX AC AAW29716;  
 XX DT 09-NOV-1998 (first entry)  
 XX DE Mouse neurotrophic factor NNT-1.  
 XX NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
 KW peripheral neuropathy; dystrophy; neural retina degeneration;  
 KW common variable immunodeficiency; CVID; selective IGA deficiency;  
 KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;  
 therapy.  
 XX Mus sp.  
 XX Key Location/Qualifiers  
 FH Key 1..27  
 FT Peptide /label= Sig\_peptide  
 FT Protein 28..225  
 FT /label= Mat\_protein  
 XX WO9833922-A1.  
 XX 06-AUG-1998.

XX 02-FEB-1998; 98WO-US02363.  
 XX 30-JAN-1998; 98US-0016534.  
 PR 03-FEB-1997; 97US-0792019.  
 XX (AMGE-) AMGEN INC.  
 XX Chang M, Elliot GS, Sarmiento U, Senaldi G;  
 PI WPI; 1998-437475/37.  
 DR N-PSDB; AAV47512.  
 XX Newly isolated nucleic acid encoding human or murine neurotrophic  
 PT factor NNT-1 - useful for treatment of neurological and  
 PT immunological diseases or inflammation, also as vaccine adjuvant  
 XX Claim 13; Fig 5; 120pp; English.  
 XX This is the amino acid sequence of a murine neurotrophic factor,  
 CC designated NNT-1, that is a growth factor for neurons and for B or  
 CC T cells. It was deduced from isolated NNT-1 cDNA (see AAV47512).  
 CC Human NNT-1 (see AAW29715) is also provided. Vectors and host cells  
 CC for use in the production of human murine recombinant NNT-1  
 CC polypeptides. These are used to treat: (i) neurological or  
 CC immunological diseases, specifically Alzheimer's, Parkinson's  
 CC or Huntington's diseases, amyotrophic lateral sclerosis,  
 CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and  
 CC degeneration of the neural retina, or conditions characterised by T  
 CC or B cell defects, e.g. common variable immunodeficiency (CVID),  
 CC selective IGA deficiency, hypogammaglobulinaemia and X-linked  
 CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)  
 CC antibody production following vaccination, and, since it inhibits  
 CC tumour necrosis factor production, it may also be useful for  
 CC treating sepsis. In addition, cells that have been engineered to  
 CC express NNT-1 can be implanted, or nucleic acids are delivered in  
 CC gene therapy vectors.  
 XX SQ Sequence 225 AA;  
 Alignment Scores:  
 Pred. No.: 1-62e-104 Length: 225  
 Score: 1224.00 Matches: 225  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 81.71% Indels: 0  
 DB: 19 Gaps: 0  
 US-09-931-704-4 (1-819) x AAW29716 (1-225)  
 QY 95 ATGGACCTCCGAGCGGACCTGCTGGGGATGTTAGCTTGCATGCGGTGCTGTGG 154  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 155 CACCTCCCTGCGAGTCCAGCTCTTAATCCACAGGAGATCCAGGCCCTGCGCCCTCCATC 214  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
 QY 215 CAGAAACCTATGACTCACTACCCGCTACCTGGAGCATCACTCCGCGAGCTTAGCTGGGACC 274  
 Db 41 GlnLysThrThrAspLeuThrArgTrpLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 275 TACCTGAATCTACCTGGGCGCCCTTTCAACGAGCCCTGACTTCAATCTCTCGACTGGGG 334  
 Db 61 TyrLeuAsnTyrLeuGlyProProAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 80  
 QY 335 GCAGAACTCTGCCAGCGCCAGCTCACTTGGAGTGTGGCAAGCCCTCAATGACAGG 394  
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100  
 QY 395 CTGGCGGTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTGTACTTGGTGGCCNC 454

Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 455 AACCTCAGGCTGCCACAGCTGAATCCAGCTAGCTGGCCCACTTCTGTACAGGCTC 514  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 515 CAGGCGCTGCTGGCAGCATTGTCAGGTGTCATGCGCAGCTTGGCTACCCACTGCCCCAG 574  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160  
 QY 575 CCTCTGCCAGGACTGAGCCAGCCTGGGCGCCCTGGCCCTGCCACAGTGTGCTTCTCCAG 634  
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180  
 QY 635 AAGATGGATGACTTCTGCTGCTGAGGAGCTGCAGACCTGGCTATGGCGTTGAGCCCAAG 694  
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 695 GACTTCAACCGCTTAAGAAGATGCGAGCTCCAGCAGCTTCAGTCCACCTGCACCTTG 754  
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
 QY 755 GAGGCACATGGTTTC 769  
 Db 221 GluAlaHisGlyPhe 225  
 RESULT 3  
 AAW56142  
 ID AAW56142 standard; Protein; 225 AA.  
 AC AAW56142;  
 XX 13-JUL-1998 (first entry)  
 DT Amino acid sequence of murine neurotrophic factor NNT-1.  
 DE Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;  
 DE treatment; neurological disease; degeneration; Parkinson's disease;  
 DE amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.  
 XX Mus sp.  
 Key Location/Qualifiers  
 FT Peptide 1..27  
 FT Protein /note= "signal peptide"  
 FT 28..225  
 FT /note= "mature peptide"  
 XX US5741772-A.  
 PN 21-APR-1998.  
 PD 03-FEB-1997; 97US-0792019.  
 XX 03-FEB-1997; 97US-0792019.  
 PR (AMGE-) AMGEN INC.  
 XX Chang M;  
 PI WPI; 1998-260526/23.  
 DR N-PSDB; AAV22654.  
 XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -  
 PT useful for stimulating growth of motor and sympathetic neurons  
 XX Claim 2; Fig 5; 41pp; English.  
 XX The present sequence represents a murine neurotrophic factor, designated  
 CC NNT-1, which is capable of stimulating growth of motor or sympathetic  
 CC neurons. The NNT-1 protein is useful in the treatment of neurological  
 CC diseases characterised by the degeneration and death of particular  
 CC classes of neurons. These diseases specifically include Parkinson's

CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,  
 CC stroke and various degenerative disorders affecting vision.

XX  
 SQ Sequence 225 AA;

Alignment Scores:  
 Pred. No.: 1,62e-104 Length: 225  
 Score: 1224.00 Matches: 225  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 81.71% Indels: 0  
 DB: 19 Gaps: 0

US-09-931-704-4 (1-819) x AAW56142 (1-225)

QY 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCCTATGACGGTGTGTGG 154  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 155 CACCTCCCTGCGAGTCCAGCTCTTAATCCACAGGAGATCCAGGCCCTGCCCTCCATC 214  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
 QY 215 CAGAAACCTATGACCTCACCCCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274  
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 275 TACCTGAATACCTGGGGCCCTTTCAAGAGCCCTGACTCAATCTCTCGACTGGG 334  
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 QY 335 CGAGAACTCTGCCAGGCGCAGGTCACTTGGAGTGTGGAGCCCTCAATGACAGG 394  
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100  
 QY 395 CTGCGGCTGACCCAGCAATATGAGCGGTACAGTCACTCTGTGTACTTTCGCGTGCCTC 454  
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 455 ACCGTGAGCTGCCAGCTGAACTCCGAGTAGCTGCGCCCACTTCTGTACCAAGCTC 514  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 515 CAGGGCTGCTGGCGCAGCATTCAGGTGTCTATGGCGACGCTTGGCTACCCACTGCCCGAG 574  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160  
 QY 575 CTTCTGCCAGGAGTACAGCAGCTGGGCGCCCTGGCCCTGCCCACAGTACTTCTCCAG 634  
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180  
 QY 635 AAGATGATGACTCTCGCTGCTGAAGAGCTGCAGACCTGCTATGCGTTTCAGCCCAAG 694  
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 695 GACTTCAACCGGCTTAAGAAGATGACGAGCTCCAGCAGCTTTCAGTCCACCTGCACTTG 754  
 Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
 QY 755 GAGGCATCATGTTTC 769  
 Db 221 GluAlaHisGlyPhe 225

RESULT 4  
 ID AAY87814 standard; Protein; 225 AA.  
 XX  
 AC  
 XX  
 XX  
 DT 24-AUG-2000 (first entry)  
 XX  
 DE Murine NNT-1 protein.  
 XX

KW NNT-1; neurotrophic factor; neurotrophic; neuroprotective; treatment;  
 KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;  
 KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; murine;  
 KW Huntington's disease; peripheral neuropathy; neural retina degeneration;  
 KW retinopathy; immune disorder; hematopoietic disorder.

XX Mus sp.

XX US6054294-A.

XX 25-APR-2000.

XX 12-DEC-1997; 97US-0988819.

XX 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.

XX Chang M;

XX WPI; 2000-338492/29.

XX N-PSDB; AAA39483.

XX New nucleic acids encoding neurotrophic factors useful for stimulating  
 growth of motor or sympathetic neurons for treating neuron cell damage

Claim 2b; Fig 5; 42pp; English.

This invention describes a novel nucleic acid molecule (I) encoding a  
 novel neurotrophic factor (NNT-1) (II) which has neurotrophic,  
 neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and  
 ophthalmological activity. (I) is useful for producing NNT-1  
 polypeptides which are useful for treating patients in whom various  
 cells of the central, autonomic, or peripheral nervous system have  
 degenerated and/or have been damaged by congenital disease, trauma,  
 mechanical damage, surgery, stroke, ischemia, infection, metabolic  
 disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1  
 proteins are used to treat diseases like Alzheimer's, Parkinson's,  
 amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's  
 disease, peripheral neuropathy or degeneration of the neural retina such  
 as retinitis pigmentosa, drug-induced retinopathies, stationary forms of  
 night blindness, progressive cone-rod degeneration, immune disorders and  
 hematopoietic disorders. (I) is effective in treating neurological  
 conditions and promotes neuron regeneration. Neural functions are  
 effectively restored in patients suffering from various neurological  
 disorders. This sequence represents the murine NNT-1 protein described in  
 the method of the invention.

SQ Sequence 225 AA;

Alignment Scores:  
 Pred. No.: 1,62e-104 Length: 225  
 Score: 1224.00 Matches: 225  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 81.71% Indels: 0  
 DB: 21 Gaps: 0

US-09-931-704-4 (1-819) x AAY87814 (1-225)

QY 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCCTATGACGGTGTGTGG 154  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 155 CACCTCCCTGCGAGTCCAGCTCTTAATCCACAGGAGATCCAGGCCCTGCCCTCCATC 214  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
 QY 215 CAGAAACCTATGACCTCACCCCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274

Db 41. GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 275 TACCTGAACCTACCTGGGCCCCCTTTCAACAGAGCTGACTTCAATCCCTCCGACTGGG 334  
 Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProPheAsnGly 80  
 QY 335 GCAGAACTCTGCCAGGCGCAGCTCACTTGAAGTGTGGCAGGCTCAATGACAGG 394  
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTyrArgSerLeuAsnAspArg 100  
 QY 395 CTGGGCTGACCCAGAACTATGAGCGTACAGTCACTCCCTGCTGTACTTGGTGGCTC 454  
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 455 AACCGTCAGGCTGCCACAGCTCACTGAGCTCGAGCTGAGCTGGCCCACTTCTGACAGCTC 514  
 Db 121 AsnArgGlnAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 515 CAGGCTGCTGGGAGCAGCTTGCAGTGTCTATGGCGAGCTTGGCTACCCACTGCCCCAG 574  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160  
 QY 575 CCTCTGCAGGACTGAGCAGCTGGGCCCCCTGGCCCTGCGCTGAGCTGAGCTGAGCTGAG 634  
 Db 161 ProLeuProGlyThrGluProAlaTyrPalapProGlyProAlaHisSerAspPheLeuGln 180  
 QY 635 AAGATGATGACTTCTGGCTGCTGAAGAGCTGCAGAGCTGCGCTATGGCTGAGCTGAGCTGAG 694  
 Db 181 LysMetAspAspPheTyrLeuLeuLysGluLeuGlnThrTyrPleuTyrArgSerAlaLys 200  
 QY 695 GACTTCAACCGCTTAAAGAGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 754  
 Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
 QY 755 GAGGCACATGGTTTC 769  
 Db 221 GluAlaHisGlyPhe 225

RESULT 5  
 AAU78177  
 ID AAU78177 standard; Protein; 225 AA.  
 XX  
 AC AAU78177;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE Mouse novel neurotrophic factor NNT1.  
 XX  
 KW Mouse; NNT1; neurotrophic factor; Ige-related disease;  
 KW Type 1 allergic disease; allergic rhinitis; eczema; dermatitis;  
 KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;  
 KW vascular stenosis; rheumatoid arthritis; psoriatic arthritis;  
 KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;  
 KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;  
 KW inflammatory bowel disease; transplant rejection; reproductive disorder;  
 KW graft versus host disease; infertility; miscarriage; preterm labour.  
 XX  
 OS Mus sp.  
 XX  
 PN WO200215977-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 17-AUG-2001; 2001WO-US25906.  
 XX  
 PR 18-AUG-2000; 2000US-226436P.  
 PR 16-AUG-2001; 2001US-0931704.  
 XX  
 PA (AMGE-) AMGEN INC.  
 XX  
 PI Senaldi G;  
 XX  
 DR WPI; 2002-280867/32.

DR N-PSDB; ABK11649.

XX Treating Immunoglobulin E-related disease, modulating Ige levels in a  
 PT patient, preventing Ige-related disease and treating allergic diseases,  
 PT involves administering NNT-1 inhibitor to a patient -

XX Claim 2; Fig 5; 63pp; English.

XX The invention relates to treating Immunoglobulin E (Ige)-related disease,  
 CC modulating Ige levels in a patient, preventing an Ige-related disease,  
 CC and treating allergic diseases, comprising administering a  
 CC therapeutically effective amount of novel neurotrophic factor (NNT)-1  
 CC inhibitor to a patient. Also included are a method of diagnosing an  
 CC Ige-related disease or susceptibility to an Ige-related disease, by  
 CC determining the presence or amount of expression of an NNT1 polypeptide  
 CC encoded by a NNT1 nucleotide sequence, its fragment or naturally  
 CC occurring variant, and diagnosing an Ige-related disease or  
 CC susceptibility of an Ige-related disease based on the presence or amount  
 CC of expression of the polypeptide and a pharmaceutical composition for use  
 CC in treating Ige-related disease, comprising the NNT1 inhibitor.  
 CC The NNT1 inhibitor is useful for preventing and treating Ige-related  
 CC disease, modulating Ige levels, and treating allergic diseases e.g.  
 CC Type 1 allergic disease, allergic rhinitis, eczema, dermatitis,  
 CC pollinosis, asthma, immune diseases and disorders, diseases involving  
 CC abnormal cell proliferation including cancer, arteriosclerosis and  
 CC vascular stenosis, diseases and conditions relating to dysfunction of  
 CC immune system including rheumatoid arthritis, psoriatic arthritis,  
 CC inflammatory arthritis, osteoarthritis, inflammatory joint disease,  
 CC autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory  
 CC bowel disease, transplant rejection, and graft versus host disease, and  
 CC reproductive diseases and disorders including infertility, miscarriage,  
 CC preterm labour and delivery, and endometriosis. The present sequence  
 CC represents Mouse NNT1.

XX SQ Sequence 225 AA;

Alignment Scores:

Pred. No.: 1-62e-104 Length: 225  
 Score: 1224.00 Matches: 225  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 81.71% Indels: 0  
 DB: 23 Gaps: 0

US-09-931-704-4 (1-819) x AAU78177 (1-225)

QY 95 ATGGACTCTGACGAGGAGCTCTGGGGATGTTAGCTTCCCTATGACGGTCTGTGG 154  
 Db 1 MetAspLeuArgAlaGlyAspSerTyrGlyMetLeuAlaCysLeuCysThrValLeuTyr 20  
 QY 155 CACCTCCCTGCGAGTCCAGCTCTTAATCCACAGGAGATCCAGGCCCTGCCCCCTCATC 214  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
 QY 215 CAGAAACCTATGACCTCACTCCCGCTACCTGGAGCATCACTCCGAGCTTAGCTGGGACC 274  
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 275 TACCTGAACCTACCTGGGCCCCCTTTCAACAGAGCTGACTTCAATCCCTCGACTGGG 334  
 Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProPheAsnGly 80  
 QY 335 GCAGAACTCTGCCAGGCGCAGCTCACTTGAAGTGTGGCAGGCTCAATGACAGG 394  
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTyrArgSerLeuAsnAspArg 100  
 QY 395 CTGGGCTGACCCAGAACTATGAGCGTACAGTCACTCCCTGCTGTACTTGGTGGCTC 454  
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 455 AACCGTCAGGCTGCCACAGCTCACTGAGCTCGAGCTGAGCTGGCCCACTTCTGACAGCTC 514  
 Db 121 AsnArgGlnAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140





```

Db 221 GlyAlaHisGlyPhe 225
|||||
RESULT 7
AAW56141
ID AAW56141 standard; Protein; 225 AA.
AC AAW56141;
XX
XX 13-JUL-1998 (first entry)
XX
XX Amino acid sequence of human neurotrophic factor NNT-1.
XX
XX Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
XX treatment; neurological disease; degeneration; Parkinson's disease;
XX amyotrophic lateral sclerosis; ALS; Alzheimers's disease; stroke.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..27
XX Protein /note= "signal peptide"
XX /note= "mature protein"
XX
XX US5741772-A.
XX
XX 21-APR-1998.
XX
XX 03-FEB-1997; 97US-0792019.
XX
XX 03-FEB-1997; 97US-0792019.
XX
XX (AMGE-) AMGEN INC.
XX
XX Chang M;
XX
XX WPI; 1998-260526/23.
XX N-PSDB; AAV22852.
XX
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
XX useful for stimulating growth of motor and sympathetic neurons
XX
XX Claim 1; Fig 3; 41pp; English.
XX
XX The present sequence represents a human neurotrophic factor, designated
XX NNT-1, which is capable of stimulating growth of motor or sympathetic
XX neurons. The NNT-1 protein is useful in the treatment of neurological
XX diseases characterised by the degeneration and death of particular
XX classes of neurons. These diseases specifically include Parkinson's
XX disease, amyotrophic lateral sclerosis (ALS), Alzheimers's disease,
XX stroke and various degenerative disorders affecting vision.
XX
XX Sequence 225 AA;

Alignment Scores:
Pred. No.: 1.18e-101 Length: 225
Score: 1193.00 Matches: 218
Percent Similarity: 98.22% Conservative: 3
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 79.64% Indels: 0
DB: 19 Gaps: 0

US-09-931-704-4 (1-819) x AAW56141 (1-225)

Qy 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCTATCCAGCGTGTGG 154
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
Qy 155 CACCTCCCTGCGAGTCCAGCTTTAATCCACAGGAGATCCAGGCCCTGCCCTCCATC 214
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 215 CAGAAAACCTATGACCTCACCGCTACCTGGAGCATCAACTCCCGAGCTTAGCTGGGACC 274
Db 41 GlnLysThrTrpAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 275 TACCTGAACCTACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCTCTCTCGACTGGGG 334
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
QY 335 GCAGAAACTCTGCCAGGGCCACGGTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGG 394
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
QY 395 CTGCGGCTGACCCAGAACTATGAGCGGTACAGTCACCTCTGCTGTACTTCCGTGGGCTC 454
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 455 AACCGTCCAGGTGCCACAGCTGAACCTCCGACGTAGCTGGCCCTTCTGTACCGAGCTC 514
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
QY 515 CAGGGCTGCTGGGAGCATTCGAGGTGTATGCGGACGCTTGCTTACCCACCTGCCAG 574
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
QY 575 CCTCTGCCAGGAGTACGACGAGCTGGGCCCTCGGCTGCTGCCACAGTACTTCTCCAG 634
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
QY 635 AAGATGGATGACTTCTGGCTGTGAGAGAGCTGCAGACCTTGGCTATGGCGTTTCCAGCAAG 694
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 695 GACTTCAACCGGCTTAAGAAGAGATGCGCTCCAGCAGCTTCAGTACCCCTGCATTTG 754
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
QY 755 GAGGCACATGGTTTC 769
Db 221 GlyAlaHisGlyPhe 225

RESULT 8
AAW94466
ID AAW94466 standard; Protein; 225 AA.
XX
XX AAW94466;
XX
XX 22-APR-1999 (first entry)
XX
XX Human cardiotrophin-like cytokine protein.
XX
XX Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;
XX CLC; IL-6; diagnosis; detection; immune system-related disorder;
XX cancer; cardiac disorder; heart failure; hypertension; cancer;
XX autoimmune disorder; infection.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..27
XX Protein /label= signal
XX /label= 28..225
XX Domain /label= Cardiotrophin-like_cytokine
XX /label= 74..79
XX /note= CD-I
XX /note= "conserved domain"
XX Domain 150..156
XX /label= CD-II
XX /note= "conserved domain"
XX Domain 194..198
XX /label= CD-III
XX /note= "conserved domain"
XX
XX WO9900415-A1.

```

XX 07-JAN-1999.  
XX 29-JUN-1998; 98WO-US13129.  
XX 30-JUN-1997; 97US-0051311.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Ruben SM, Shi Y;  
XX WPI; 1999-095678/08.  
XX N-PSDB; AAX16161.  
XX New isolated cardiostrophin-like cytokine nucleic acid - used to  
XX develop products for treating cardiac and immune system disorders,  
XX e.g. heart failure, hypertension, cancers, autoimmune disorders and  
XX infections  
XX Claim 1; Fig 1; 103pp; English.  
XX The present invention relates to a novel cardiostrophin-like cytokine  
XX (CLC) protein which is a member of the interleukin 6 (IL-6) cytokine  
XX family. The present sequence represents the human CLC protein. The  
XX present invention also describes screening methods for identifying  
XX agonists and antagonists of CLC activity, as well as methods for  
XX detecting cardiac and immune system-related disorders and  
XX therapeutic methods for treating cardiac and immune system-related  
XX disorders, e.g. heart failure, hypertension, cancers, autoimmune  
XX disorders and infections.  
XX SQ Sequence 225 AA;  
Alignment Scores:  
Pred. No.: 1,18e-101 Length: 225  
Score: 1193.00 Matches: 218  
Percent Similarity: 98.22% Conservative: 3  
Best Local Similarity: 96.89% Mismatches: 4  
Query Match: 79.64% Indels: 0  
DB: 20 Gaps: 0  
US-09-931-704-4 (1-819) x AAW94466 (1-225)  
QY 95 ATGACCTCCGAGCGGGGCTGCTGGGGATGTAGCTTGCCTATGCACGGTGTCTGG 154  
DB 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuGlyThrValLeuTrp 20  
QY 155 CACCTCCCTGCGAGTCCAGCTCTTAATCGCAGAGATCCAGGCGCTGGCCCTCCATC 214  
DB 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40  
QY 215 CAGAAAACCTATGACCTACCCGCTACCTGAGGATCACTCCGAGCTTAGCTGGACC 274  
DB 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
QY 275 TACCTGACTACCTGGGGCCCTTTCAACGAGCTGACTTCAATCCTCTCGACTGGGG 334  
DB 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
QY 335 GCAGAAACTCTGCCAGGGCCACCGCTCAACTTGGAGTGTGGCGAAGCTCAATGACAGG 394  
DB 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
QY 395 CTGGGGTGCACCAACTATGAGCGGTACAGTCACTCCCTGTGTACTTGGGTGGCCCTC 454  
DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuLeuLeuGlyLeu 120  
QY 455 AACCGTACGGTGGCCACAGCTGAATCCGACGTAGCTGGCCCACTTCTGTACCGACCTC 514  
DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
QY 515 CAGGCGCTCTGCTGGGAGCATTCAGGTGTATGGCGAGCGCTTGGCTTACCCACTGCCCCAG 574

DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
QY 575 CCTCTGCCAGGACTGAGCCAGCTGGCCCTGGCCCTGCCACAGTACTTCTCCAG 634  
DB 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
QY 635 AAGATGATGACTTCTGGCTGCTGAAGAGCTGCAGAGCTGGCTATGGCGTTCACCCAG 694  
DB 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
QY 695 GACTTCAACCGGCTTAAGAGAGATGAGCTCCAGCAGCTTCAGTCCACCTGCACCTG 754  
DB 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaAlaValThrLeuHisLeu 220  
QY 755 GAGGCACATGGTTC 769  
DB 221 GlyAlaHisGlyPhe 225  
RESULT 9  
ID AAY87813 standard; Protein; 225 AA.  
XX AAY87813;  
XX 24-AUG-2000 (first entry)  
XX Human NNT-1 protein.  
XX NNT-1; human; neurotrophic factor; neurotrophic; neuroprotective; treatment;  
XX anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;  
XX nervous system degeneration; Alzheimer's disease; Parkinson's disease;  
XX amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
XX Huntington's disease; peripheral neuropathy; neural retina degeneration;  
XX retinopathy; immune disorder; hematopoietic disorder.  
XX Homo sapiens.  
XX US6054294-A.  
XX 25-APR-2000.  
XX 12-DEC-1997; 97US-0988819.  
XX 03-FEB-1997; 97US-0792019.  
XX (AMGE-) AMGEN INC.  
XX Chang M;  
XX WPI; 2000-338492/29.  
XX N-PSDB; AAA39481.  
XX New nucleic acids encoding neurotrophic factors useful for stimulating  
XX growth of motor or sympathetic neurons for treating neuron cell damage  
Claim 1c; Fig 3; 42pp; English.  
XX This invention describes a novel nucleic acid molecule (I) encoding a  
XX novel neurotrophic factor (NNT-1) (II) which has neurotrophic,  
XX neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and  
XX ophthalmological activity. (I) is useful for producing NNT-1  
XX polypeptides which are useful for treating patients in whom various  
XX cells of the central, autonomic, or peripheral nervous system have  
XX degenerated and/or have been damaged by congenital disease, trauma,  
XX mechanical damage, surgery, stroke, ischemia, infection, metabolic  
XX disease, nutritional deficiency, malignancy and/or toxic agents, NNT-1  
XX proteins are used to treat diseases like Alzheimer's, Parkinson's,  
XX amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's  
XX disease, peripheral neuropathy induced by diabetes or other metabolic  
XX disorders, and/or dystrophies or degeneration of the neural retina such  
XX as retinitis pigmentosa, drug-induced retinopathies, stationary forms of  
XX night blindness, progressive cone-rod degeneration, immune disorders and

CC hematopoietic disorders. (I) is effective in treating neurological  
 CC conditions and promotes neuron regeneration. Neural functions are  
 CC effectively restored in patients suffering from various neurological  
 CC disorders. This sequence represents the human NNT-1 protein described in  
 CC the method of the invention.

XX Sequence 225 AA;

Alignment Scores:  
 Pred. No.: 1.18e-101 Length: 225  
 Score: 1193.00 Matches: 218  
 Percent Similarity: 98.22% Conservatives: 3  
 Best Local Similarity: 96.89% Mismatches: 4  
 Query Match: 79.64% Indels: 0  
 DB: 21 Gaps: 0

US-09-931-704-4 (1-819) x AAY87813 (1-225)

QY 95 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTAGCTTGCTATGACGGTGTGTGG 154  
 Db |||||  
 1 MecAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 155 CACCTCCTCGAGCGCCAGCTCTTAATCGCACAGAGATCCAGGCCCTGCCCTCCATC 214  
 Db |||||  
 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
 QY 215 CAGAAACCTATGACCTACCCGCTACCTGAGCATCAACTCCGAGCTTAGCTGGGACC 274  
 Db |||||  
 41 GlnLysThrTrpAspLeuThrArgTrpLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 275 TACCTGAACCTACCTGGGGCCCTTTCAACGAGCGCTGACTCAATCTCTCGACTGGGG 334  
 Db |||||  
 61 TyrLeuAsnTrpLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 QY 335 GCAGAACTCTGCCAGGGCCACGGTCAACTGGAAGTGTGGCGAAGCTCAATGACAGG 394  
 Db |||||  
 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
 QY 395 CTGGGCTGACCCAGAACTATGAGGCGTACAGTCACTCTCTGTGTACTTCTCGTGGCCTC 454  
 Db |||||  
 101 LeuArgLeuThrGlnAsnTrpGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 455 AACCGTCAGGCTGCCAGCTGAACTCCGAGCTAGCTGGCCGCTCTGTACAGGCTC 514  
 Db |||||  
 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 515 CAGGGCTCTGGGAGCATTTGAGGTGTCTATGGCGACGCTTGCTACCCACTGCCCCAG 574  
 Db |||||  
 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
 QY 575 CCTCTGCCAGGAGCTAGCCAGCTGGGCCCTGGGCCCTGCCACAGTACTTCTCTCCAG 634  
 Db |||||  
 161 ProLeuProGlyThrGluProTrpTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
 QY 635 AAGATGATGACTCTGGCTGTGAAGGAGCTGCAGACCTGGCTATGGCTTACGCCAAG 694  
 Db |||||  
 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnTrpLeuTrpArgSerAlaLys 200  
 QY 695 GACTCAACCGCTTAAAGAGATGACGCTCCAGCAGCTTCAGTCACTGCACTGCTG 754  
 Db |||||  
 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
 QY 755 GAGGCACATGTTTC 769  
 Db |||||  
 221 GlyAlaHisGlyPhe 225

RESULT 10  
 AAG63543  
 ID AAG63543 standard; Protein; 225 AA.  
 XX  
 AC AAG63543;  
 XX

DT 15-OCT-2001 (first entry)

XX

DE Amino acid sequence of a human NNT-1 protein.

XX

KW NNT-1; CLF-1; sCNTFRalpha; nervous system; neuron; nervous system;  
 KW neuro-muscular function; tumour; immune system; haematopoietic system;  
 KW reproductive system; liver; skeletal muscle; neurodegenerative disease;  
 KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;  
 KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;  
 KW blastocyst implantation; thrombosis; retinal disease;  
 KW retinal pigmentosis.

XX Homo sapiens.

XX WO200155172-A2.

XX 02-AUG-2001.

XX 26-JAN-2001; 2001WO-FR00253.

XX 27-JAN-2000; 2000FR-0001035.

XX 12-OCT-2000; 2000FR-0013089.

XX (FABR) FABRE MEDICAMENT SA PIERRE.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;

XX WPI; 2001-488773/53.

XX N-PSDB; AAH74484.

XX A complex comprising a NNT-1 protein and a CLF-1 and/or sCNTFRalpha  
 protein useful to treat neurodegenerative disease including Parkinson's  
 and Huntington's, obesity and cancer -

XX Claim 2; Page 58; 67pp; French.

CC The present sequence represents a human NNT-1 protein. The specification  
 CC describes a complex comprising a NNT-1 protein and a CLF-1 and/or  
 CC sCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate  
 CC activity of the sCNTFRalpha/gp130/LiFrbeta receptor complex, or to  
 CC induce phosphorylation of the tyrosine of gp130 and LiFrbeta,  
 CC particularly where cells expressing the receptor complex are in the  
 CC central or peripheral nervous system, in neurons implicated in  
 CC neuro-muscular function or in skeletal muscle. The complex or  
 CC antibodies are also used to decrease the survival, growth or  
 CC proliferation of tumour cells or to facilitate the proliferation and/or  
 CC inhibit differentiation of cells stocks. The complex is also used to  
 CC modulate activity of the gp130/LiFrbeta receptor or cells expressing  
 CC that receptor, particularly those cells implicated in the immune,  
 CC haematopoietic, nervous or reproductive system, the liver or skeletal  
 CC muscle. Molecules of the invention may be used to prevent or treat  
 CC neurodegenerative diseases including amyotrophic lateral sclerosis,  
 CC Parkinson's and Huntington's disease, to repair or regenerate nervous  
 CC or muscular tissue or to maintain muscular mass in paralysis patients.  
 CC They may also be used to treat cancer, obesity and associated diseases,  
 CC and to improve fertility, particularly to avoid endometriosis and/or  
 CC assist blastocyst implantation, thrombosis, or retinal disease,  
 CC particular retinal pigmentosis.

XX Sequence 225 AA;

Alignment Scores:

Pred. No.: 1.18e-101 Length: 225  
 Score: 1193.00 Matches: 218  
 Percent Similarity: 98.22% Conservatives: 3  
 Best Local Similarity: 96.89% Mismatches: 4  
 Query Match: 79.64% Indels: 0  
 DB: 22 Gaps: 0

US-09-931-704-4 (1-819) x AAG63543 (1-225)

QY 95 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTAGCTTGCTATGACGGTGTGTGG 154  
 |||||

1	MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp	20
155	CACCTCCCTGCAGTGCAGCTCTTAATCGCACAGAGATCCAGGCCCTGCCTCCATC	214
21	HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle	40
215	CAGAAACCTATGACCTCACCCGCTACCTGGAGCATCAACTCCCACTTAGCTGGGACC	274
41	GlnLysThrTyAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr	60
275	TACCTGAACACTACCTGGGGCCCCCTTCAACAGAGCTGACTTCAATCTCTCCAGCTGGG	334
61	TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly	80
335	GCAGAAACTCTGCCAGGGCCACAGGTCAACTTGGAAAGTGTGGCGAAGCCTCAATGACAGG	394
81	AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys	100
395	CTGGCGGTACCCAGAACTATGAGCGGTACAGTCACTCCCTGTTGTTACTTCGGTGGCTC	454
101	LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu	120
455	AACCGTCAGCGTCCACAGCTCAACTCCGACGCTAGCCTGGGCCCACTTCTGTACCAAGCTC	514
121	AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu	140
515	CAGGGCTCTCGGGCAGCAATTCAGGGTGTATGGCGACGCTTGGCTACCCACTGCCCCAG	574
141	GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaLeuAlaLeuGlyTyrProLeuProGln	160
575	CTCTGCCAGGACGTAGCAGCGCTGGGCCCTGGCCCTGCCACAGTACGATTCCTCCAG	634
161	ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln	180
635	AAGATGGATGACTTTCGTGCTGTAAGGAGCTGCAGACCTGCTAGTATGGCGTTACGCCAAG	694
181	LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys	200
695	GACTTCAACCGGCTTAAGAGAGATGACGCTCCAGCAGCTTCAGTCAACCTGCACCTTG	754
201	AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu	220
755	GAGGCACATGGTTTC	769
221	GlyAlaHisGlyPhe	225
RESULT 11		
AAU78176		
ID	AAU78176 standard; Protein; 225 AA.	
XX		
AC	AAU78176;	
XX		
DT	05-JUN-2002 (first entry)	
XX		
DE	Human novel neurotrophic factor NNT1.	
XX		
KW	Human; NNT1; neurotrophic factor; IgE-related disease;	
KW	Type I allergic disease; allergic rhinitis; eczema; dermatitis;	
KW	pollinosis; asthma; immune disease; cancer; arteriosclerosis;	
KW	vascular stenosis; rheumatoid arthritis; psoriatic arthritis;	
KW	inflammatory arthritis; osteoarthritis; inflammatory joint disease;	
KW	autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;	
KW	inflammatory bowel disease; transplant rejection; reproductive disorder;	
KW	graft versus host disease; infertility; miscarriage; preterm labour.	
XX		
OS	Homo sapiens.	
XX		
PN	W0200215977-A2.	
XX		
PD	28-FEB-2002.	
XX		
PF	17-AUG-2001; 2001WO-US25906.	
XX		



QY 301 CAACAGCGCTGATCTCAATCTCTCGACTGGGGGAGAACTCTCCAGGCGCCACGGT 360  
Db 165 eAsnGluProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrVa 185  
QY 361 CAACCTGGAAGTGTGGGAGGCTCAATGACAGGCTGGCGGTGACCCAGAACTATGAGGC 420  
Db 185 lAspLeuGluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAl 205  
QY 421 GTACAGTCACTCTGTGTACTTGGTGGCTCTCAACCGCTCAGGCTGCCACAGCTGAAC 480  
Db 205 aTyrSerHisLeuLeuGlyCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLe 225  
QY 481 CCGAGTACGCTGGCGCACTCTGTACAGGCTCCAGGCTCTGCCAGGACTCAGCCAGCTG 540  
Db 225 uArgSerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGl 245  
QY 541 TGCTATGGGAGCGCTGGCTTACCACTGCCCCAGCTCTGCCAGGACTCAGCCAGCTG 600  
Db 245 yValMetAlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTr 265  
QY 601 GGCCCCCTGGCCCTGCCACAGTCACTTCTCCAGAGATGATGATCTTGGCTGTGAA 660  
Db 265 pThrProGlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLy 285  
QY 661 GGAGTGCAGACTGCTGATGGCTTACAGCAAGGACTTCAACCGCTTAAAGAGAGAT 720  
Db 285 sGluLeuGlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMe 305  
QY 721 CGAGCTCCAGCAGCTTCAGTCACTCCCTGCACTTGGAGGCACATGTTTC 769  
Db 305 tGlnProProAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 321  
RESULT 13  
ID AAM79399 standard; Protein; 321 AA.  
XX AC AAM79399;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human protein SEQ ID NO 3045.  
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX OS Homo sapiens.  
XX PN WO200157190-A2.  
XX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US04098.  
XX PR 03-FEB-2000; 2000US-0496914.  
XX PR 27-APR-2000; 2000US-0560875.  
XX PR 20-JUN-2000; 2000US-0598075.  
XX PR 19-JUL-2000; 2000US-0620325.  
XX PR 01-SEP-2000; 2000US-0654936.  
XX PR 15-SEP-2000; 2000US-0663561.  
XX PR 20-OCT-2000; 2000US-0693325.  
XX PR 30-NOV-2000; 2000US-0728422.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52532.

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX Claim 20; Page 237; 6221pp; English.  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX SQ Sequence 321 AA;  
Alignment Scores: 1.61e-101 Length: 321  
Pred. No.: 1192.00 Matches: 226  
Percent Similarity: 89.49% Conservative: 4  
Best Local Similarity: 87.94% Mismatches: 14  
Query Match: 79.57% Indels: 13  
DB: 22 Gaps: 3  
US-09-931-704-4 (1-819) x AAM79399 (1-321)  
QY 37 CCCACTCCGCCAGCTCTGGGAGA---GGAGCCGGGCCCC---GGCCGGCCCCCCCCCAG 90  
Db 66 ProHisProProSerProArgTrpGlyGlnThrProGluGlyLeuProAlaAlaSer 85  
QY 91 CCCCATGGACCT-----CCGAGCAGGCGACTCTG 120  
Db 86 ProCysGlyProGlyProArgSerCysPheSerSerIleLeuProThr-GlyAspSerTr 105  
QY 121 GGGGATGTAGCTTGCCTATGCACGCTGTGGACCTCCCTGCAGTGCAGCTCTTAA 180  
Db 105 pGlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAs 125  
QY 181 TCGCAGCAGAGATCCAGGCGCTTCCATCCAGAAACCTATGACCTCACCCTA 240  
Db 125 nArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTy 145  
QY 241 CCTGAGCATCAATCCCGAGCTTAGCTGGGACCTACCTGAACTACTGCGGCCCCCTTT 300  
Db 145 rLeuGluHisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPh 165  
QY 301 CAACGAGCTGACTTCAATCTCTCGACTGGGGGAGAACTCTGCCAGGCGCCACGGT 360  
Db 165 eAsnGluProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrVa 185  
QY 361 CAACCTGGAAGTGTGGGAGGCTCAATGACAGGCTGGCGGTGACCCAGAACTATGAGGC 420  
Db 185 lAspLeuGluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAl 205  
QY 421 GTACAGTCACTCTGTGTACTTGGTGGCTCTCAACCGCTCAGGCTGCCACAGCTGAAC 480  
Db 205 aTyrSerHisLeuLeuGlyCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLe 225  
QY 481 CCGAGTACGCTGGCGCACTTCTGTATACAGCTCTGCCAGGCTCTGCGGCGCATTTGAGG 540  
Db 225 uArgSerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGl 245  
QY 541 TGCTATGGGAGCGCTGGCTTACCACTGCCCCAGCTCTGCCAGGACTCAGCCAGCTG 600  
Db 245 yValMetAlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTr 265



XX Human; biologically active complex; haemopoietin receptor; NR6;  
 KW cardiotrophin-like cytokine; CLC; therapy; prophylaxis; proliferation;  
 KW differentiation; cell survival; neurotrophic activity.  
 XX Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT Peptide 1..27  
 FT Protein /label= Signal\_peptide  
 FT 28..223  
 FT /label= Human mature CLC protein  
 FT /note= "Cardiotrophin-like cytokine"  
 XX WO200127157-A1.  
 XX 19-APR-2001.  
 XX 06-OCT-2000; 2000WO-AU01216.  
 XX 08-OCT-1999; 99AU-0003327.  
 XX 12-MAY-2000; 2000AU-0007489.  
 XX (AMRA-) AMRAD OPERATIONS PTY LTD.  
 XX Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;  
 PI Nakata Y, Hasegawa M;  
 PI N-PSDB; AAD04201.  
 DR WPI; 2001-281978/29.  
 DR N-PSDB; AAD04201.  
 XX New biologically active complex comprising NR6 and  
 PT cardiotrophin-like-cytokine, for facilitating proliferation,  
 PT differentiation and/or survival of a cell -  
 XX Claim 32; Page 114-115; 123pp; English.  
 XX The present invention relates to a biologically active complex comprising  
 CC a haemopoietin receptor, NR6 and cardiotrophin-like cytokine (CLC).  
 CC The complex is useful in the manufacture of a medicament for the  
 CC treatment and/or prophylaxis of a subject, as it is involved in  
 CC facilitating proliferation, differentiation and/or survival of a cell.  
 CC The complex or its components have neurotrophic activity. The present  
 CC sequence is human cardiotrophin-like cytokine (CLC) protein.  
 XX  
 SQ Sequence 223 AA;  
 Alignment Scores:  
 Pred. No.: Length: 223  
 Score: 1181.00 Matches: 216  
 Percent Similarity: 98.21% Conservative: 3  
 Best Local Similarity: 96.86% Mismatches: 4  
 Query Match: 78.84% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-931-704-4 (1-819) x AAE00828 (1-223)  
 QY 95 ATGGACCTCCGACGAGGACTCGTGGGGATGTTAGCTTGCTATGCACGGTGTGGG 154  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 QY 155 CACCTCCCTGCAGTCCAGCTCTTAAATCGACAGGAGATCCAGGCCCTGGCCCTCCATC 214  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40  
 QY 215 CAGAAACCTATGACCTCACCGGTACCTGGAGCATCACTCCGAGCTTAGCTGGGACC 274  
 Db 41 GlnIysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 QY 275 TACCTGAACCTACCTGGGCCCTTTCAACGAGCCTGACTTCAATCCTCCTCGACTGGGG 334  
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 335 GCAGAACTCTGCCAGGGCCACCGTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGG 394  
 Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
 QY 395 CTGCGGTGACCCAGAACTATGAGCGGTACAGTCACCTCCTGTGTACTTGGTGGCCCTC 454  
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 QY 455 AACCGTCCAGGCTGCCACAGCTGAACCTCCGACGTAGCTGGCCCACTTCTGTACCAAGCCTC 514  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 QY 515 CAGGCCCTGTGGCAGCATTGCAGGTGTATGGCGAGCGCTTGGCTACCCACTGCCCCAG 574  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
 QY 575 CCTCTGCCAGGACTGAGCCAGCCTGGGCCCTGGCCCTGCCACAGTCACTTCTCTCCAG 634  
 Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
 QY 635 AAGATGGATGACTTCTGCTGCTGAAGGAGCTGCAGACTTGGCTATGGCGTTCAGCCAG 694  
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 QY 695 GACTTCAACCGGCTTAAAGAAGATGCAGCCTCCAGCAGCTTCAGTCACCTGCACCTTG 754  
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
 QY 755 GAGGCACAT 763  
 Db 221 GlyAlaHis 223

Search completed: January 27, 2003, 15:59:53  
 Job time : 38.4559 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 16:20:23 ; Search time 5.86484 Seconds  
(without alignments)  
5635.701 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 1498

Sequence: 1 tattattaaagcttcgcgg.....aggccagtcagctgtgctt 819

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 244452

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09931704/runat\_27012003\_154129\_3760/app\_query.fasta\_1.7189  
-DB=Published\_applications\_AA -QFM=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US0931704 @CGN 1 1 51 @runat\_27012003\_154129\_3760  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:\*\*

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1224	81.7	225	10	US-09-931-704-5
2	1193	79.6	225	10	US-09-931-704-2
3	852	56.9	164	10	US-09-864-761-40014
4	134	9.2	529	10	US-09-861-597-2

Sequence	2,	Appli
Sequence 2,	Appli	
Sequence 50,	Appli	
Sequence 47,	Appli	
Sequence 56,	Appli	
Sequence 4,	Appli	
Sequence 2,	Appli	
Sequence 17,	Appli	
Sequence 217,	Appli	
Sequence 217,	Appli	
Sequence 1,	Appli	
Sequence 70,	Appli	
Sequence 2,	Appli	
Sequence 8,	Appli	
Sequence 10,	Appli	
Sequence 2,	Appli	
Sequence 5,	Appli	
Sequence 19,	Appli	
Sequence 720,	Appli	
Sequence 138,	Appli	
Sequence 979,	Appli	
Sequence 14,	Appli	
Sequence 24,	Appli	
Sequence 1002,	Appli	
Sequence 3,	Appli	
Sequence 4,	Appli	
Sequence 3,	Appli	
Sequence 2,	Appli	
Sequence 9,	Appli	
Sequence 2,	Appli	
Sequence 216,	Appli	
Sequence 216,	Appli	
Sequence 34988,	Appli	
Sequence 36844,	Appli	
Sequence 3,	Appli	
Sequence 164,	Appli	
Sequence 254,	Appli	
Sequence 1422,	Appli	
Sequence 2,	Appli	

#### ALIGNMENTS

RESULT 1  
US-09-931-704-5  
; Sequence 5, Application US/09931704  
; Patent No. US20020041873A1  
; GENERAL INFORMATION:  
; APPLICANT: Senaldi, Giorgio  
; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using Inhibitors  
; TITLE OF INVENTION: Inhibitors  
; FILE OF INVENTION: A-695  
; CURRENT APPLICATION NUMBER: US/09/931,704  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: US 60/226,436  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 225  
; TYPE: PPT  
; ORGANISM: Murine  
; ORGANISM: Murine  
US-09-931-704-5

Pred. No.:	Length:
4,54e-90	225
1224.00	225
Score:	Matches:
Percent Similarity:	Conservative:
Best Local Similarity:	Mismatches:
Query Match:	Indels:
DB:	Gaps:

US-09-931-704-4 (1-819) x US-09-931-704-5 (1-225)

QY 95 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTTGCCTATGACGGTGTGTGG 154  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20

QY 155 CACCTCCCTGACGTGCACCTCTTAATCGCACAGGAGATCCAGGCCCTGCGCCCTCCATC 214  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 215 CAGAAAACCTATGACTCAACCGCTACCTGGAGCATCAACTCCGCGAGCTTAGCTGGGACC 274  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 275 TACCTGAATACCTGGGGCCCTTTCAACGAGCCCTGACTTCAATCCTCCCTCCAGCTGGGG 334  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 335 CGAGAAACTCTGCCAGGGCCACGGTCAACTTGGAAAGTGTGGCGAAGCCCTCAATGACAGG 394  
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100

QY 395 CTGGGGCTGACCCAGAACATATAGGCGTACAGTCACCTCTCTGTGTACTTGTGCGCTC 454  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120

QY 455 RACCGCTCAGGCTGCCACAGCTGAACCTCCGACGTAGCTGGCCCTCTCTGTACAGCCTC 514  
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 515 CAGGGCTCTGGGCGACGATTCAGAGTGTCTATGGCGACGCTGGCCCTCTCTGTACAGCCTC 574  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160

QY 575 CTTCTGCCAGGAGTACGACGCTGGGCGCTGGGCGCTGGCCCTGCCACAGTACTTCTCCAG 634  
Db 161 ProLeuProGlyThrGluProAlaTyrAlaProGlyProAlaHisSerAspPheLeuGln 180

QY 635 AGATGGAGTACTTCTGGCTGTCTGAGGAGCTGCAGACCTGGCTATGGCGTTCAGCCCAAG 694  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACCGGCTTAAGAGAGATGACGCTCCAGCAGCTTCAGTCACCCCTGCACCTG 754  
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220

QY 755 GAGGCACATGGTTTC 769  
Db 221 GluAlaHisGlyPhe 225

## RESULT 2

US-09-931-704-2  
; Sequence 2, Application US/09931704  
; Patent No. US20020041873A1  
; GENERAL INFORMATION:

; APPLICANT: Senaldi, Giorgio  
; TITLE OF INVENTION: Methods and Compositions for Treating Igb-Related Disease Using N  
; TITLE OF INVENTION: Inhibitors  
; FILE REFERENCE: A-695  
; CURRENT APPLICATION NUMBER: US/09/931,704  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: US 60/226,436  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 225  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-931-704-2

## Alignment Scores:

Pred. No.: 1,33e-87 Length: 225  
Score: 1193.00 Matches: 218  
Percent Similarity: 98.22% Conservative: 3  
Best Local Similarity: 96.89% Mismatches: 4  
Query Match: 79.64% Indels: 0  
DB: 10 Gaps: 0

US-09-931-704-4 (1-819) x US-09-931-704-2 (1-225)

QY 95 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTTGCCTATGACGGTGTGTGG 154  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20

QY 155 CACCTCCCTGACGTGCACCTCTTAATCGCACAGGAGATCCAGGCCCTGCGCCCTCCATC 214  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 215 CAGAAAACCTATGACTCAACCGCTACCTGGAGCATCAACTCCGCGAGCTTAGCTGGGACC 274  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 275 TACCTGAATACCTGGGGCCCTTTCAACGAGCCCTGACTTCAATCCTCCCTCCAGCTGGGG 334  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 335 CGAGAAACTCTGCCAGGGCCACGGTCAACTTGGAAAGTGTGGCGAAGCCCTCAATGACAGG 394  
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100

QY 395 CTGGGGCTGACCCAGAACATATAGGCGTACAGTCACCTCTCTGTGTACTTGTGCGCTC 454  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120

QY 455 AACCGTCCAGGCTGCCACAGCTGAACCTCCGACGTAGCTGGCCCTCTCTGTACAGCCTC 514  
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 515 CAGGGCTCTGGGCGACGATTCAGAGTGTCTATGGCGACGCTGGCCCTCTCTGTACAGCCTC 574  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaLeuGlyTyrProLeuProGln 160

QY 575 CTTCTGCCAGGAGTACGACGCTGGGCGCTGGGCGCTGGCCCTGCCACAGTACTTCTCCAG 634  
Db 161 ProLeuProGlyThrGluProTrpTrpThrProGlyProAlaHisSerAspPheLeuGln 180

QY 635 AAGATGATGACTTCTGGCTGTCTGAAAGAGCTGCAGACCTGGCTATGGCGTTCAGCCCAAG 694  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACCGGCTTAAGAGAGATGACGCTCCAGCAGCTTCAGTCACCCCTGCACCTG 754  
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220

QY 755 GAGGCACATGGTTTC 769  
Db 221 GlyAlaHisGlyPhe 225

## RESULT 3

US-09-864-761-40014  
; Sequence 40014, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aesomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04

;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 40014  
;; LENGTH: 164  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC005849.1  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5  
;; OTHER INFORMATION: EST HUMAN HIT: AI752561.1, EVALUATE 3.00e-66  
;; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUATE 8.00e-03  
US-09-864-761-40014

Alignment Scores:  
Pred. No.: 1.68e-60 Length: 164  
Score: 852.00 Matches: 156  
Percent Similarity: 97.56% Conservative: 4  
Best Local Similarity: 95.12% Mismatches: 4  
Query Match: 56.88% Indels: 0  
DB: 10 Gaps: 0

US-09-931-704-4 (1-819) x US-09-864-761-40014 (1-164)

QY 278 CTGAACCTACCTGGGCCCCCTTCAACGAGCTGACTTCAATCTCTCGACTGGGGCA 337  
Db 1 LeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProPheAsnGlyAla 20  
QY 338 GAACTCTGCCAGGGCCCGCTCAACTTGGAGTGTGGCGAAGCTCAATGACAGGCTG 397  
Db 21 LysThrLeuProArgAlaThrValAspLeuGluValTyrArgSerLeuAsnAspLysLeu 40  
QY 398 CGGCTGACCCAGAACTATGAGCGGTACAGTACCTCTCTGTGTACTTGGTGGCCTCAAC 457

Db 41 ArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsn 60  
QY 458 CGTCAGGCTGCCACAGCTCAACTCCGACCTAGCTGGCGCCACTTCTGTACCGCTCCAG 517  
Db 61 ArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeuGln 80  
QY 518 GGCCTGCTGGCAGCATTCAGGTGTATGGGAGCGCTTGGCTTACCACTGCCCCAGCCT 577  
Db 81 GlyLeuLeuGlySerIleAlaGlyValMetAlaLeuGlyTyrProLeuProGlnPro 100  
QY 578 CTGCCAGGACTGAGCCAGCCTGGGCCCTGGCCCTGCCACAGTCACTTCTCTCCAGAG 637  
Db 101 LeuProGlyThrGluProThrTyrProGlyProAlaHisSerAspPheLeuGlnLys 120  
QY 638 ATGGATGACTTCTGCTGCTGAAGGAGTGCAGACCTGGCTATGGCGTTCAGCAGGAC 697  
Db 121 MetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLysAsp 140  
QY 698 TTCAACCGGCTTAAAGAAGATGCAGCCTCCAGCAGCTTCACTACCTGCACCTTGGAG 757  
Db 141 PheAsnArgLeuLysLysLysMetGlnProProAlaAlaAlaValThrLeuHisLeuGly 160  
QY 758 GCACATGGTTTC 769  
Db 161 AlaHisGlyPhe 164

## RESULT 4

US-09-861-597-2  
; Sequence 2, Application US/09861597  
; Patent No. US20020064539A1  
; GENERAL INFORMATION:  
; APPLICANT: PHILLIPPE, Michel  
; APPLICANT: GARSON, Jean-Claude  
; APPLICANT: ARRAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; TITLE OF INVENTION: ANALOG  
; FILE REFERENCE: 6388-0365-0  
; CURRENT APPLICATION NUMBER: US/09/861,597  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: 09/247,806  
; PRIOR FILING DATE: 1999-02-11  
; PRIOR APPLICATION NUMBER: FR 98/01614  
; PRIOR FILING DATE: 1998-02-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 529  
; TYPE: PRT  
; ORGANISM: Nephila clavipes  
US-09-861-597-2

Alignment Scores:  
Pred. No.: 0.00266 Length: 529  
Score: 134.00 Matches: 74  
Percent Similarity: 38.08% Conservative: 17  
Best Local Similarity: 30.96% Mismatches: 97  
Query Match: 9.17% Indels: 51  
DB: 10 Gaps: 12

US-09-931-704-4 (1-819) x US-09-861-597-2 (1-529)

QY 671 TCTGCAGCTCTTCCAGCAGCCAGAGTCAATCTTCTGGAGGAGTCACTC----- 621  
Db 73 SerAlaAlaAlaAlaAlaAlaGlySerGlyGlnGlnGlyProGlyGlyTyrGlyProArg 92  
QY 620 -----TGTGGCAGGCGC---CAGGGGCCAGGCTGGCTCAGTCCCTG 582  
Db 93 GlnGlnGlyProGlyGlyTyrGlyGlnGlnGlyProSerGlyProGlySerAla 112  
QY 581 GCAGAGGCTGGGGCAGTGGGTAGCCAAAGCGTGCATGACACCTCAATGTCTGCCAGCA 522

```
Db 113 AlaAlaIaSerAlaAlaSerAlaGluSer-----Gly 124
Qy 521 GGCCCTGGAGGCTGTACAGAGTGGCCAGGCTACGTCCGAGTTTCAGCTGTGGCAGCCT 462
Db 125 GlyProGlyGlyTyr-----GlyProGlyGlnGlnGlyProGlyGlyTyrGlyPro 141
Qy 461 GACGGTTGAGGCACGCAAGTAACACAGAGGAGTGACTGTACGCCTCATATGTTCTGGGTCA 402
Db 142 GlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProSerGlyProGlySer 161
Qy 401 GCCCAGGCTGTATGAGGCTTCCACACACTTCCAAAGTTGACCGTGGCCCTGGCAGAG 342
Db 162 AlaAlaAlaAlaAlaAlaAlaAlaAlaSerGlyProGlyGlnGlnGlyProGlyGln 179
Qy 341 TTTCTGCCCCAGTCCGA-----GGAGGATTGAAGTCAGGCTCGTTG---AAAGGGG 294
Db 179 TyrGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProSerG 199
Qy 293 GCCCAGGTTAGTTCAGGTAGTCCAGCTAAGCTGC-----GGAGTTGATGCTCCA 243
Db 199 lyProGlySerAlaAlaAlaAlaAlaAlaAlaAlaSerGlyProGlyGlnGlnGlyPro 218
Qy 242 GGTAGC---GGGTGAGTTCATAGTTTCTGGATGAGGGGCCAGGCGCTGGATCTCCTG 186
Db 219 GlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGly----- 234
Qy 185 TCGGATTAAAGCTGCGCACTGCGAGGAGTGGCCACAGCACCGTGCATAGGCAAGCTAACA 126
Db 235 -----GlnGlnGlyLeuSerGlyProGlySerAlaAlaAlaAla 247
Qy 125 TCCCCACAGTCCCCTGCTCGAGTCCATGGGGCTGGGGCGCGCGCGCGCGCG 66
Db 248 AlaAlaAlaGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyPro 267
Qy 65 GTCCTCTCCAGAGCTGGCGAGTGGGAGGCGGAGCGCGCTCGCGG 15
Db 268 Ser-----GlyProGlySerAlaAlaAlaAlaAlaAla 278

RESULT 5
US-09-823-240-2
; Sequence 2, Application US/09823240
; Patent No. US20020048813A1
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehlend
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; FILE REFERENCE: Motility
; CURRENT APPLICATION NUMBER: US/09/823,240
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-823-240-2

Alignment Scores:
Pred. No.: 0.00783 Length: 802
Score: 128.50 Matches: 73
Percent Similarity: 39.82% Conservative: 17
Best local Similarity: 32.30% Mismatches: 82
Query Match: 8.58% Indels: 54
DB: 10 Gaps: 15

US-09-931-704-4 (1-819) x US-09-823-240-2 (1-802)
Qy 34 CCTCCCACTCCGCCA-----GCCTCTGGGAGAGAGCGCGCGCGCGGCG---CGGCCC 81
```

```
Db 381 ProProSerProProIleMetIleSerSerProProGlyLysAlaThrGlyProArgPro 400
Qy 82 -----GGCCCCCAGCCCATCGACCTCCGAGC----- 108
Db 401 ValLeuProValCysValSerSerProValProGlnMetProProSerProThrAlaPro 420
Qy 109 AGGGGACTCGTGGGGATGTAGCTGCCTATG-----CAGGTCTCTGTG 153
Db 421 AsnGlySerLeuAaspSerValThrTyrProValSerProProThrSerGlyProAla 440
Qy 154 GCACCTCCCTCCAGTGCAGCTCTTAATCGCACAGGAGATCCAGGCGCTGGCCCTCCAT 213
Db 441 AlaProProProProProPro-----ProProProProPro 453
Qy 214 CCAGAAAACCTATGACCTCACCCCTACCTGGA---GCATCAACTCCCGCAGCTTAGCTGG 270
Db 454 ProProProLeuProProProProProProProLeuAlaSerLeuSerHisCysGlySer 473
Qy 271 GACCTACCTGAACCTACCTGGGGCCCCCTTT-----CAACGAGCCTGA 312
Db 474 GlnAlaSerProProProGlyThrProLeuAlaSerThrProSerSerLysProSerVal 493
Qy 313 CTTCAATCCTCTCAGTCTGGGGGAGAACTCTGCCAGGCG-----CACGGTCAACTT 366
Db 494 LeuProSerProSerAlaGlyAlaProAlaSerAlaGluThrProLeuAsnProGluLeu 513
Qy 367 GGA-----AGTGTGGCGAGCCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
Db 514 GlyAspSerSerAlaSerGluProGlyLeuGlnAlaAlaSerGlnPro----- 529
Qy 421 GTACAGTACCT---CCTGTGTACTTGTGGTGGCTCAACCGTCAAGCTGAGGTGCCACAGCTGA 477
Db 530 AlaGluSerProThrProGlnGlyLeuValLeuGlyProProAla----- 544
Qy 478 ACTCCGAGTAGCTTGGCCCACTTCTGTACAGGCTCCAGGCGCTGGGCGAGCATTCG 537
Db 545 ---ProProProProProProLeu-----ProSerGlyProAla-TyrAlaSerAl 560
Qy 538 AGGTGTCATGGCGAGCTTGGCTACCCACTGCCCGAGCTCTGCCAGGAGCTGAGCCAGC 597
Db 560 a---LeuProProProProGlyProProProProProProProProProProPro 579
Qy 598 CTGGGCGCTGGCCCT 613
Db 579 oProProProProPro 584

RESULT 6
US-10-001-873-50
; Sequence 50, Application US/10001873
; Patent No. US20020160388A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and
; FILE REFERENCE: DEX-0275
; CURRENT APPLICATION NUMBER: US/10/001,873
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,055
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,496
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 1134
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-873-50
```





```

      |||||||
Db 891 ArgGlyAlaArgGlyProThrGlyLysProGlyProLysGlyThrSerGlyGlyAspGly 910
QY 91 CCCCATGGACCTCCGAGC-----AGGGACTCGTGGGGGATGTAGCTTCCT 138
      |||||||
Db 911 ProProGlyProProGlyGluArgGlyProGlnGlyProGlnGlyProValGlyPhePro 930
QY 139 ATGCAGGTGCTGTGCACCTCCCTCAGTGCACGT----- 175
      |||||||
Db 931 -----GlyProLysGlyPro-ProGlyProProGlyArgMetGlyCysProGlyHisPr 948
QY 176 -----CTTAATGCACAGGAGATCCAGGCCCTCGCCCTC 210
Db 948 oGlyGlnArgGlyGluThrGlyPheGlnGlyThrGlyProProGlyProGlyGlyVa 968
QY 211 CATC-----CAGAAACCTATGACCTCACCTGCTACCTGAGCATCACTCCGAGCTT 264
      |||||||
Db 968 lValGlyProGlnGlyProThrGlyGluThrGlyProLleGlyGluArg----- 984
QY 265 AGCTGGGACCTACCTGAACTACTTGGGGCCCTTTCACAGAGCCTGACTTCAATCTCTCC 324
      |||||||
Db 985 -----GlyTyProGlyProPro-----GlyProPr 993
QY 325 T-----CGACTGGGGGCGAGAACTCTGCCAGGGC 354
Db 993 oGlyGlnGlnGlyLeuProGlyAlaAlaGlyLysGlyGlyAla-LysGlyAspProGlyP 1013
QY 355 CACGCTCA---ACTTGAAGTGTGGGAAGCCTCAATGACAGGCTGC---GGCTGACCCA 408
      |||||||
Db 1013 roGlnGlyLeuSerGlyLysAspGlyProAlaGlyLeuArgGlyPheProGly-GluArg 1032
QY 409 GAACATCAGGGGTACAGTCACTCTGTGTACTTGTGGCTGGCTCAACGTCAGGCTGC 468
      |||||||
Db 1033 GlyLeuProGlyAlaGlnGlyAlaProGlyLeuLysGlyGlyGluGlyProGlnGly--- 1051
QY 469 CACAGCTGAACCTCCGAGCTAGCTGSCCCACTCTGTACACAGCTCCAGGGCTGTGGG 528
      |||||||
Db 1052 -----ProProGlyPro-ValGl 1057
QY 529 CAGC-----ATTGAGGTGTATGGCGAGCGTTGGCTACCC 564
      |||||||
Db 1057 ySerProGlyGluArgGlySerAlaGlyThrAlaGlyProLleGlyLeuArgGlyArgPr 1077
QY 565 ACTGCCACGCTCTGCCAGGAGCTAGCCAGC-----TGGGCCCT 607
      |||||||
Db 1077 oGlyProGlnGlyProProGly-----ProAlaGlyGluLysGlyAlaProGlyGly 1095
QY 608 -----GGCCTGCGCCACAGTGACTTCTCCAG 634
      |||||||
Db 1095 sGlyProGlnGlyProAlaGlyArgAspGlyValGln 1107

RESULT 10
US-09-770-689A-4
; Sequence 4, Application US/09770689A
; Patent No. US2002011517A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001079
; CURRENT APPLICATION NUMBER: US/09/770,689A
; CURRENT FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 726
; TYPE: PRT
; ORGANISM: HUMAN
US-09-770-689A-4
Alignment Scores:
Pred. No.: 0.0211 Length: 726
```

```

Score: 123.00 Matches: 54
Percent Similarity: 36.63% Conservative: 20
Best Local Similarity: 26.73% Mismatches: 72
Query Match: 8.21% Indels: 56
DB: 10 Gaps: 12

US-09-931-704-4 (1-819) x US-09-770-689A-4 (1-726)
QY 31 CGCCTCCCTCCACTCCGCCAGCCTCTGGGAGAGAGCGCGCGCGCGGCC---GGCCCC 87
      |||||||
Db 518 ArgSerProSerProProThrGlnHisThrGlyGlnProProGlyGlnProSerAlaPro 537
QY 88 CAGCCCCATGAGACCTCCGAGCAGG-----GGACTCTGTGGGGGATGTAGCTTGCCTATG 141
      |||||||
Db 538 SerGlnLeuSerAlaProArgArgTySerSerSerLeuSerProIleGlnAlaProAsn 557
QY 142 CACGGTCTGTGGCACCTCCCTCGAGTCCAGTCTTAAATCCACAGGAGTCCAGG--- 198
      |||||||
Db 558 HisProProGlnProProThrGlnAlaThrProLeuMetHisThrLysProAsnSer 577
QY 199 -----CCTCGGCCCTC-----CATCCAGAAAC 222
      |||||||
Db 578 GlnGlyProProAsnProMetAlaLeuProSerGluHisGlyLeuGluGlnProSerHis 597
QY 223 CTATGACCTCACCCGCTACTCTGGAGCATCAACTCCGACGTTAGCTGGGACCTACCTGAA 282
      |||||||
Db 598 ThrProProGlnThrProThrProProSerThrProProLeuGlyLysGlnAsnProSer 617
QY 283 CTACTCGGGGCCCTTTCACAGGAGCTGACTTCAATCTCTCGACTGGGGGCGAGAAC 342
      |||||||
Db 618 LeuPro---AlaPro-----GlnThrLeuAlaGlyGlyAsnProGlu 630
QY 343 TCTGCCAGGGCCACGGTCAACTTGAAGTGTGGGAAGCCTCAATGACAGCTGGCGCT 402
      |||||||
Db 631 ThrAlaGlnProHisAla-----GlyThrLeuProArgProArg----- 643
QY 403 GACCCAGAACTATGAGCGGTACATCCTCTGTGTACTTGTGGCTGACCGCTCA 462
      |||||||
Db 644 -----ProVal-----ProLysProArg 649
QY 463 GGCTGCCACAGCTCAACTCCGACGCTAGCTGGCCACCTTCTGTACCAGCTCCAGGGCT 522
      |||||||
Db 650 AsnArgProSerValProPro-----ProProGlnPro 660
QY 523 GCTGGGAGCATTCGAGGTGTCTATGCGAGCGCTTGGCTACCC---ACTGCCCGCCCTCT 579
      |||||||
Db 661 ProGlyValHisSerAla-----GlyAspSerSerLeuThrAsnThrAlaProThrAla 678
QY 580 GCCAGG 585
      |||||||
Db 679 SerLys 680

RESULT 11
US-09-770-689A-2
; Sequence 2, Application US/09770689A
; Patent No. US2002011517A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001079
; CURRENT APPLICATION NUMBER: US/09/770,689A
; CURRENT FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 803
; TYPE: PRT
; ORGANISM: HUMAN
US-09-770-689A-2
Alignment Scores:
```

Pred. No.: 0.0214 Length: 803  
Score: 123.00 Matches: 54  
Percent Similarity: 36.63% Conservative: 20  
Best Local Similarity: 26.73% Mismatches: 72  
Query Match: 8.21% Indels: 56  
DB: 10 Gaps: 12

US-09-931-704-4 (1-819) x US-09-770-689A-2 (1-803)

```
QY 31 CGCCCTCCCACTCCGCGCAGCTCTGGGAGAGGAGCGCGCGCGCGGCC---GGCCCC 87
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 ArgSerProSerProThrGlnHisThrGlnProProGlyGlnProSerAlaPro 614
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 88 CAGCCCCATGACCTCCGAGCAGG---GGACTCGTGGGGAGTGTAGCTTGCCTATG 141
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 615 SerGlnLeuSerAlaProArgArgTyrSerSerSerLeuSerProIleGlnAlaProAsn 634
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 142 CACGGTGTGTGGCACCTCCCTGCAGTGCAGCTCTTAATCGCACAGGATCCAGG--- 198
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 HisProProGlnProProThrGlnAlaThrProLeuMetHisThrLysProAsnSer 654
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 199 -----CCCTGGCCCTC-----CATCAGAAAC 222
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 GlnGlyProProAsnProMetAlaLeuProSerGluHisGlyLeuGlnProSerHis 674
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 223 CTATGACCTCACCGCTACCTGGAGCATCAACTCCGAGCTTAGCTGGGACCTACCTGAA 282
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 ThrProProGlnThrProThrProProSerThrProProLeuGlyLysGlnAsnProSer 694
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 283 CTACCTGGGGCCCCCTTTCAACAGAGCTGACCTCAATCCTCCTCGACTGGGGGACAGAAAC 342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 695 LeuPro---AlaPro-----GlnThrLeuAlaGlyGlyAsnProGlu 707
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 343 TCTGCCAGGGCCAGCTCACTTGAAGTGTGGGAGAGCTCAATGACAGGCTGGGCT 402
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 708 ThrAlaGlnProHisAla-----GlyThrLeuProArgProArg----- 720
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 403 GACCAGAACTAGAGCGGTACAGTCACTCTCTGTGTACTTGGCGGCTCAACCGTCA 462
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 -----ProVal-----ProLysProArg 726
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 463 GGCTGCCACAGCTGAATCCGAGTAGCTGGCCCACTTCTGTACCAGCTCCAGGGCCT 522
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 AsnArgProSerValProPro-----ProGlnPro 737
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 523 GCTGGGAGCATTCAGGTGTATGGCGAGCGTGGCTACCC---ACTGCCCGAGCCTCT 579
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 738 ProGlyValHisSerAla-----GlyAspSerSerLeuThrAsnThrAlaProThrAla 755
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 580 GCCAGG 585
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 756 SerLys 757
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 12

```
US-09-816-860A-2
; Sequence 2, Application US/09816860A
; Patent No. US20020081651A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 26649, A No. US20020081651A1 Human GTPase Activating Molecule a
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: MNT-133
; CURRENT APPLICATION NUMBER: US/09/816, 860A
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/191, 859
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 881
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-860A-2
```

# Alignment Scores:

Pred. No.: 0.0218 Length: 881  
Score: 123.00 Matches: 54  
Percent Similarity: 36.63% Conservative: 20  
Best Local Similarity: 26.73% Mismatches: 72  
Query Match: 8.21% Indels: 56  
DB: 10 Gaps: 12

US-09-931-704-4 (1-819) x US-09-816-860A-2 (1-881)

```
QY 31 CGCCCTCCCACTCCGCGCAGCTCTGGGAGAGGAGCGCGCGCGGCC---GGCCCC 87
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 673 ArgSerProSerProThrGlnHisThrGlnProProGlyGlnProSerAlaPro 692
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 88 CAGCCCCATGACCTCCGAGCAGG---GGACTCGTGGGGAGTGTAGCTTGCCTATG 141
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 SerGlnLeuSerAlaProArgArgTyrSerSerSerLeuSerProIleGlnAlaProAsn 712
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 142 CACGGTGTGTGGCACCTCCCTGCAGTGCAGCTCTTAATCGCACAGGATCCAGG--- 198
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 713 HisProProGlnProProThrGlnAlaThrProLeuMetHisThrLysProAsnSer 732
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 199 -----CCCTGGCCCTC-----CATCAGAAAC 222
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 733 GlnGlyProProAsnProMetAlaLeuProSerGluHisGlyLeuGlnProSerHis 752
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 223 CTATGACCTCACCGCTACCTGGAGCATCAACTCCGAGCTTAGCTGGGACCTACCTGAA 282
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 753 ThrProProGlnThrProThrProProSerThrProProLeuGlyLysGlnAsnProSer 772
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 283 CTACCTGGGGCCCCCTTTCAACAGAGCTGACCTCAATCCTCCTCGACTGGGGGACAGAAAC 342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 773 LeuPro---AlaPro-----GlnThrLeuAlaGlyGlyAsnProGlu 785
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 343 TCTGCCAGGGCCAGCTCACTTGAAGTGTGGGAGAGCTCAATGACAGGCTGGGCT 402
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 786 ThrAlaGlnProHisAla-----GlyThrLeuProArgProArg----- 798
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 403 GACCAGAACTAGAGCGGTACAGTCACTCTCTGTGTACTTGGCGGCTCAACCGTCA 462
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 799 -----ProVal-----ProLysProArg 804
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 463 GGCTGCCACAGCTGAATCCGAGTAGCTGGCCCACTTCTGTACCAGCTCCAGGGCCT 522
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 805 AsnArgProSerValProPro-----ProGlnPro 815
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 523 GCTGGGAGCATTCAGGTGTATGGCGAGCGTGGCTACCC---ACTGCCCGAGCCTCT 579
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 816 ProGlyValHisSerAla-----GlyAspSerSerLeuThrAsnThrAlaProThrAla 833
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 580 GCCAGG 585
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 834 SerLys 835
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 13

```
US-10-086-464-17
; Sequence 17, Application US/10086464
; Publication No. US20020199218A1
; GENERAL INFORMATION:
; APPLICANT: GORING, Daphne R. et al.
; TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES
; FILE REFERENCE: P 25,762-A USA
; CURRENT APPLICATION NUMBER: US/10/086,464
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 10/069,304
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: PCT/CA00/00966
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/149,466
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: US 60/159,122
; PRIOR FILING DATE: 1999-10-13
```





```
Db 200 GlnAlaLeuProLeuLeuLeuArgGln***HisCys***ProProSerTyrSerIle 219
Qy 284 TACCTGGGGCCCTTTCAACGAGCTGACTTCAATCCTCTCGAC----- 329
Db 220 TyrLeuArgProSerLeu-ThrGlyValArgGluValArgGlyAsp***LeuSerGlySe 239
Qy 330 -----TGGGGGCGCAAACTC-----TGCCAGGG---CCACGGTCAACTTGGAA 370
Db 239 rArgProTrpMetProGlyPheProAlaGlyCysProAlaCysProSer***ThrGlyLy 259
Qy 371 GTGTGGC-----GAAGCCTCAATACACAGCTGGCG 400
Db 259 sCysGlyProCysPheTrpSerCysLeuGlyThrThrArgSerAlaProThrGlyCysSe 279
Qy 401 CTGACCCAGAACTATGAGCGGTACAGTCACCTCCTGTGTTACTTGGTGGCCTCAACCGT 460
Db 279 rSerArgArgThrAlaArg-----CysGluLe 288
Qy 461 CAGGTGCGCACAGCTGAACCTCGAGCTAGCCTGGCCCACTTCTGTACCA-----GCCTC 514
Db 288 uArgSerProGln-----GlnProValSerValProGlyArgSerPr 302
Qy 515 CAGGCGCTGCTGGGAGCAGATTGCAGGTGTATGGGAGCGCTGGCTACCACTGCCCGG 574
Db 302 oArgAlaLeuTrpArgProPro-----ArgArgAsnThrAspProArgAr 318
Qy 575 CCTCTGCCAGGAGCTGAGCCAGCCTGGGCGCTGGCCCTGCCACAGTCACTTCTCCAG 634
Db 318 GluValGlnLeuLeuArgGln-----HisSerSe 328
Qy 635 AAGATGATGACTTCTGGCTGTGTAAGGAGCTGC-----AGACCTGGCTA 679
Db 328 rProTrpGlnValTyGlyPheValArgAlaCysLeuArgGluValProProGlyLe 348
Qy 680 T-----GGGTTTCAGCCAGGACTTCAAC 703
Db 348 uTrpGlySerArgHisAsnGluArgPheLeuArgAsnThrLysLysPheIleSerLe 368
Qy 704 CGGCTTAAGAGAAGATGAGCTCCAGAGCTTCACTGACCTGCACTTGGAGGCACAT 763
Db 368 uGlyLysHisAlaLysLeuSerLeuGlnGluLeu-----ThrTrpLysMetSe 384
Qy 764 GTTTCCTGAC 773
Db 384 rValArgAsp 387

RESULT 15
US-09-438-486-217
; Sequence 217, Application US/09438486
; Publication No. US20030009019A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. US20030009019A1e1 Telomerase
; NUMBER OF SEQUENCES: 223
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/09/438,486
; FILING DATE: 12-NOV-1999
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002931US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 217:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-438-486-217
```

```
Alignment Scores:
Pred. No.: 0.0321 Length: 1003
Score: 121.00 Matches: 81
Percent Similarity: 31.17% Conservative: 20
Best Local Similarity: 25.00% Mismatches: 104
Query Match: 8.08% Indels: 119
DB: 9 Gaps: 18
```

US-09-931-704-4 (1-819) x US-09-438-486-217 (1-1003)

```
Qy 23 CCGGGCTCGCCCTCCACTCCGCGAGGAGGAGCGCGCGCGCGCGCG 82
Db 110 ProGlyValAlaLeu-----ProLeuSerArgSerGlyArgProLeuGlyArg 125
Qy 83 GCGCCAGCGCCCATGGACCTCCGAGCAGGG-----GACTCGTGGGGATGTAGCTTGC 136
Db 126 GlyProGlyPro-----ThrArgAlaGlyArgLeuAspArgValThrValSerVal 143
Qy 137 CTATGCACGGTGTGTGGCACTCCCTGCACTGCA----- 172
Db 144 TrpCys-----HisLeuProAspProLysLysProProLeuTrpArgVal 159
Qy 173 -----GCTCTTAATCGCACGAGAGTCCA 196
Db 160 ArgSerLeuAlaArgAlaThrProThrHisProTrpAlaAlaSerThrArgAlaPro 179
Qy 197 GGCCCT-----GGCCCTCCATCCAGAAACC 223
Db 180 HisProHisArgGlyHisHisValLeuGlyHisAlaLeuSerProGlyValArgArgAsp 199
Qy 224 TATGACTTACCCCTACCTACCTGGAGACATCACTCCGCACTTAGCTGGGACCTACCTGAAC 283
Db 200 GlnAlaLeuProLeuLeuArgGln***HisCys***ProProSerTyrSerIle 219
Qy 284 TACCTGGGGCCCTTTCAACGAGCTGACTTCAATCCTCTCGAC----- 329
Db 220 TyrLeuArgProSerLeu-ThrGlyValArgGluValArgGlyAsp***LeuSerGlySe 239
```

330	Qy	-----TGGGGGAGAAATC-----TGCCGAGGG-----CCAGCGTCAACTTGGAA	370
239	Db	rArgProTirpMetPheProAlaGlyCysProAlaCysProSer**ThrGlyLeu	259
371	Qy	GTGTGGC-----GAAGCCTCAATCACAGGCTGGCG	400
259	Db	sCysGlyProCysPheTirpSerCysLeuGlyThrThrArgSerAlaProThrGlyCysE	279
401	Qy	CTGACCCAGAACATATGAGGGGTACAGTCACCTCTGTGTTACTGCGTGGCCTCAACCGT	460
279	Db	rSerArgArgThrAlaArg-----CysGluLeu	288
461	Qy	CAGGCTGCCACAGCTGAACTCCGACGTAGCTGGGCCACTTCTGTACCA-----GCCTC	514
288	Db	uArgSerProGln-----GlnProValSerValProGlyArgSerPr	302
515	Qy	CAGGCGCTGCTGGGCAGCATTTGCAGGTGTGATGGCAGCAGCTTGGCTACCACTGCCCGAC	574
302	Db	oArgAlaLeuTirpArgProPro-----ArgArgAsnThrAspProArgAr	318
575	Qy	CCTCTGCCAGGAGCTGAGCCAGAGCTGGGCCCTGGCCCTGCCACAGTGACTTCTCTCCAG	634
318	Db	gLeuValGlnLeuLeuArgGln-----HisSerSe	328
635	Qy	AAGATGATGACTTCTGGCTGTGCTGAGGAGCTGC-----AGACCTGGCTTA	679
328	Db	rProTirpGlnValTyrGlyPheValArgAlaCysLeuArgLeuValProGlyLeu	348
680	Qy	T-----GGCGTTACGCCAAGACTTCAAC	703
348	Db	uTirpGlySerArgHisAsnGluArgArgPheLeuArgAsnThrLysLysPheIleSerLeu	368
704	Qy	CGGCTTAAGAAGAAGATGACGCTCCAGCAGCTTCACTCACCTGCAGTCTGGAGGCACAT	763
368	Db	uGlyLysHisAlaLysLeuSerLeuGlnGluLeu-----ThrTirpLysMetSe	384
764	Qy	GGTTTCTGCAC	773
384	Db	rValArgAsp	387

Search completed: January 27, 2003, 17:23:06  
Job time : 17.8648 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 27, 2003, 15:50:02 ; Search time 8.30852 Seconds  
(without alignments)  
5800.643 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 1498

Sequence: 1 tattattaagcttcgcgg.....aggccacgtagctgtgctt 819

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2\_1/uspro spool/US09931704/runat 27012003 154127 3671/app query.fasta\_1.7189  
-DB=Issued Patents AA -QWTF=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPFCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09931704 @CGN 1 1 89 @runat 27012003 154127 3671 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WAEV TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1224	81.7	225	1	US-08-792-019B-5
2	1224	81.7	225	3	US-08-988-819-5
3	1224	81.7	225	1	US-09-016-534-5
4	1193	79.6	225	1	US-08-792-019B-2
5	1193	79.6	225	3	US-09-106-182-2
6	1193	79.6	225	4	US-08-988-819-2
7	1193	79.6	225	3	US-09-016-534-2
8	146	9.7	330	1	US-08-642-255-32
9	146	9.7	408	1	US-07-609-716-65
10	146	9.7	408	4	US-08-475-411A-65
11	146	9.7	408	4	US-08-478-029A-65
12	140.5	9.4	1064	1	US-08-642-255-62

13	139	9.3	684	1	US-08-555-669-12	Sequence 12, Appl	
14	139	9.3	684	3	US-09-073-663-12	Sequence 12, Appl	
15	136	9.1	682	1	US-08-642-255-126	Sequence 126, App	
16	136	9.1	682	1	US-08-397-633A-36	Sequence 36, Appl	
17	136	9.1	960	4	US-09-219-849-5	Sequence 5, Appl	
18	135.5	9.0	252	1	US-08-642-255-61	Sequence 61, Appl	
19	135	9.0	357	1	US-07-609-716-66	Sequence 66, Appl	
20	135	9.0	357	1	US-08-642-255-33	Sequence 33, Appl	
21	135	9.0	357	4	US-08-475-411A-66	Sequence 66, Appl	
22	135	9.0	357	4	US-08-478-029A-66	Sequence 66, Appl	
23	135	9.2	493	4	US-08-556-978B-59	Sequence 59, Appl	
24	135	9.0	762	1	US-08-642-255-120	Sequence 120, App	
25	135	9.0	762	1	US-08-397-633A-31	Sequence 31, Appl	
C	26	135	9.2	870	2	US-09-010-928B-2	Sequence 2, Appl
C	27	134	9.2	529	4	US-09-247-806-2	Sequence 2, Appl
C	28	134	9.2	595	1	US-08-425-069-4	Sequence 4, Appl
C	29	134	9.2	595	2	US-08-317-844B-4	Sequence 4, Appl
30	133.5	8.9	829	1	US-08-642-255-132	Sequence 132, App	
31	133.5	8.9	829	1	US-08-397-633A-53	Sequence 53, Appl	
32	132	8.8	761	2	US-08-707-237A-84	Sequence 84, Appl	
33	132	8.8	762	1	US-08-642-255-114	Sequence 114, App	
34	132	8.8	762	1	US-08-397-633A-26	Sequence 26, Appl	
35	131.5	8.8	504	4	US-09-219-849-3	Sequence 3, Appl	
36	131.5	8.8	561	1	US-08-642-255-52	Sequence 52, Appl	
37	131.5	8.8	720	4	US-09-219-849-4	Sequence 4, Appl	
38	131.5	8.8	777	1	US-08-642-255-53	Sequence 53, Appl	
C	39	130	8.9	1077	1	US-07-972-032-82	Sequence 82, Appl
C	40	130	8.9	1077	1	US-08-642-255-95	Sequence 95, Appl
41	129.5	8.6	417	1	US-08-175-155-69	Sequence 69, Appl	
42	129.5	8.6	417	1	US-08-477-509B-104	Sequence 104, App	
43	129.5	8.6	417	1	US-08-642-255-102	Sequence 102, App	
44	129.5	8.6	417	2	US-08-707-237A-76	Sequence 76, Appl	
45	129.5	8.6	417	3	US-08-482-085B-104	Sequence 104, App	

ALIGNMENTS

RESULT 1

US-08-792-019B-5  
; Sequence 5, Application US/08792019B  
; Patent No. 5741772  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/792,019B  
; FILING DATE: 03-FEB-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-792-019B-5

Alignment Scores:  
Pred. No.: 3.7e-108 Length: 225  
Score: 1224.00 Matches: 225  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 81.71% Indels: 0  
DB: 1 Gaps: 0

US-09-931-704-4 (1-819) x US-08-792-019B-5 (1-225)

QY 95 ATGGACCTCCGAGCGGACCTGCTGGGGATGTTAGCTTGCTATGCAGGTCGTGG 154  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuGlyValLeuTrp 20

QY 155 CACCTCCCTGCGAGTCAGCTCTTAATCGCACAGGAGATCCAGGCCCTGGCCCTCCATC 214  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 215 CAGAAACCTATGACCTCACCCGCTACCTGGAGCATCACTCCGAGCTTAGCTGGACC 274  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 275 TACCTGAACCTACTGCGGCCCTTCAACGAGCCTGACTTCAATCCTCCTGACTGGG 334  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 455 AACCGTCAGGCTGCCACAGCTGAACCTCCGAGCTGCGGCCCTCTGTATACGACCTC 514  
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 515 CAGGCCCTGCTGGGCGAGCATTCAGGCTCATGGCGAGCTTGGCTACCCCTGCCAG 574  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160

QY 575 CCTCTGCCAGGACTGAGCCAGCTGGGCCCTGCGCCCTGCCACAGTACTTCTCCAG 634  
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180

QY 635 AAGATGGATGACTTCTGCGCTGCTGAAGGAGCTGCAGACCTGGCTATGGCGTTCCAGCAAG 694  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACCGGCTTAAGAAGATGCAGCCTCCAGCAGCTTCAGTCAACCTGCACCTG 754  
Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220

QY 755 GAGGCACATGGTTTC 769  
Db 221 GluAlaHisGlyPhe 225

RESULT 2  
US-08-988-819-5  
; Sequence 5, Application US/08988819  
; Patent No. 6054294  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; STREET: ONE AMGEN CENTER DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/988,819  
FILING DATE: 12-DEC-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/792,019  
FILING DATE: 03-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: COOK, ROBERT R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-442A  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-988-819-5

Alignment Scores:  
Pred. No.: 3.7e-108 Length: 225  
Score: 1224.00 Matches: 225  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 81.71% Indels: 0  
DB: 3 Gaps: 0

US-09-931-704-4 (1-819) x US-08-988-819-5 (1-225)

QY 95 ATGGACCTCCGAGCGGACCTGCTGGGGATGTTAGCTTGCTATGCAGGTCGTGG 154  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuGlyValLeuTrp 20

QY 155 CACCTCCCTGCGAGTCAGCTCTTAATCGCACAGGAGATCCAGGCCCTGGCCCTCCATC 214  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 215 CAGAAACCTATGACCTCACCCGCTACCTGGAGCATCACTCCGAGCTTAGCTGGACC 274  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 275 TACCTGAACCTACTGCGGCCCTTCAACGAGCCTGACTTCAATCCTCCTGACTGGG 334  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 335 GCAGAAACTCTGCGGCCGAGCTCAACTTGAAGTGTGGCGAAGCTCTGTATACGACG 394  
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100

QY 395 CTGCGGCTGACCCAGAACTATGAGCGGTACAGTCACTCCTGTTGTACTTGGTGGCCTC 454  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120

QY 455 AACCGTCAGGCTGCCACAGCTGAACCTCCGAGCTGCGGCCCTCTGTATACGACCTC 514  
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 515 CAGGCCCTGCTGGGCGAGCATTCAGGCTCATGGCGAGCTTGGCTACCCCTGCCAG 574  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160

QY 575 CCTCTGCCAGGACTGAGCCAGCTGGGCCCTGCGCCCTGCCACAGTACTTCTCCAG 634  
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180

QY 635 AAGATGGATGACTTCTGCGCTGCTGAAGGAGCTGCAGACCTGGCTATGGCGTTCCAGCAAG 694  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACCGGCTTAAGAAGAGATGCAGCCTCCAGCAGCTTCAGTCAACCTGCACCTTG 754  
 |||||  
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
 |||||  
 QY 755 GAGGCACATGGTTTC 769  
 |||||  
 Db 221 GluAlaHisGlyPhe 225  
 |||||

## RESULT 3

US-09-016-534-5  
 ; Sequence 5, Application US/09016534  
 ; Patent No. 6143874  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHANG, MING-SHI  
 ; APPLICANT: ELLIOTT, GARY S.  
 ; APPLICANT: SARMIENTO, ULLA  
 ; APPLICANT: SENALDI, GIORGIO  
 ; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
 ; NUMBER OF SEQUENCES: 16  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: AMGEN INC.  
 ; STREET: ONE AMGEN CENTER  
 ; CITY: THOUSAND OAKS  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 91320  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/016,534  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/792,019  
 ; FILING DATE: 03-FEB-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: COOK, ROBERT R.  
 ; REGISTRATION NUMBER: 31,602  
 ; REFERENCE/DOCKET NUMBER: A-442B  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 225 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-09-016-534-5

Alignment Scores:  
 Pred. No.: 3,7e-108 Length: 225  
 Score: 1224.00 Matches: 225  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 81.71% Indels: 0  
 DB: 4 Gaps: 0

US-09-931-704-4 (1-819) x US-09-016-534-5 (1-225)

QY 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCTATGACCGGTGTGTGG 154  
 |||||  
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
 |||||  
 QY 155 CACCTCCCTGCAGTCCAGCTCTTAATCGACAGAGATCCAGGCCCTGGCCCTCCATC 214  
 |||||  
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40  
 |||||  
 QY 215 CAGAAACCTATGACCTCACCGCTACCTGGAGCATCAACTCCGACGCTTAGCTGGGACC 274  
 |||||  
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
 |||||

QY 275 TACCTGAACCTACCTGGGGCCCCCTTTCAAGAGCCTGACTTCAATCCTCCTGACCTGGG 334  
 |||||  
 Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80  
 |||||  
 QY 335 GCAGAAACTCTGCCAGGCCACGGTCAACTTGGAAAGTGTGGCGAAGCCTCAATGACAG 394  
 |||||  
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100  
 |||||  
 QY 395 CTGCGGCTGACCCAGCACTATGAGGCGGTACAGTCACCTCTGTGTTACTTGCGTGGCCTC 454  
 |||||  
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
 |||||  
 QY 455 AACCGTCAGGCTGCCACAGCTGAACCTCCGACGTAGCTGGCCCTCTTCTGTACACGCTC 514  
 |||||  
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
 |||||  
 QY 515 CAGGCGCTCTGGGAGCATTCAGGTGTCATGGCGACGCTTGCTGCTACCCACTGCCCCAG 574  
 |||||  
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160  
 |||||  
 QY 575 CTCTGCGCAGGAGCTGAGCCAGCCTGGCGCCCTGGCGCTGCCACAGTGACTTCTCTCCAG 634  
 |||||  
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180  
 |||||  
 QY 635 AAGATGGATGACTTCTGCTGCTGAAGGAGCTGCAGACCTGGCTATGGGGTTTCAGGCAAG 694  
 |||||  
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
 |||||  
 QY 695 GACTTCAACCGGCTTAAGAAGAGATGCAGCCTCCAGCAGCTTCAGTCAACCTGCACCTTG 754  
 |||||  
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220  
 |||||  
 QY 755 GAGGCACATGGTTTC 769  
 |||||  
 Db 221 GluAlaHisGlyPhe 225  
 |||||

## RESULT 4

US-08-792-019B-2  
 ; Sequence 2, Application US/08792019B  
 ; Patent No. 5741772  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHANG, MING-SHI  
 ; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
 ; NUMBER OF SEQUENCES: 16  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: AMGEN INC.  
 ; STREET: 1840 DEHAVILLAND DRIVE  
 ; CITY: THOUSAND OAKS  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 91320  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/792,019B  
 ; FILING DATE: 03-FEB-1997  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: COOK, ROBERT R.  
 ; REGISTRATION NUMBER: 31,602  
 ; REFERENCE/DOCKET NUMBER: A-442  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 225 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-792-019B-2

Alignment Scores:

Pred. No.: 3 22e-105 Length: 225  
Score: 1193.00 Matches: 218  
Percent Similarity: 98.22% Conservative: 3  
Best Local Similarity: 96.89% Mismatches: 4  
Query Match: 79.64% Indels: 0  
DB: 1 Gaps: 0

US-09-931-704-4 (1-819) x US-08-792-019B-2 (1-225)

Qy 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCTATGCACGGTGTGTGG 154  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
Qy 155 CACCTCCCTGCGAGTCCAGCTCTTATCCGACGAGGATCCAGCCCTGCCCTCCATC 214  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40  
Qy 215 CAGAAAACCTATGACCTCACCGCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
Qy 275 TACCTGAACCTAGCTGGGGCCCCCTTCAACGAGCGCTGACTTCAATCTCTCGACTGGGG 334  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
Qy 335 CGAGAACTCTCCGAGCGGCGCTCACTTGGAGTGTGGGAGCCCTCAATGACAGG 394  
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
Qy 395 CTGGCGCTGACCCAGAACCTATGAGGGGTACAGTCACCTCTCTGTGTTACTTGGCGCTC 454  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
Qy 455 AACCTGAGCTGCCAGAGCTGAACCTCCGACGCTAGCTGGCGCCCTTCTGTACAGCCTC 514  
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
Qy 515 CAGGCGCTGCTGGGAGCATTCAGGTGTCATGGCGGACCTGGCTTACCCACTGCCCCAG 574  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
Qy 575 CCTCTCCAGGAGCTGAGCCAGCTGGGGCCCCCTGGCCCTGCCACAGTACTTCTCCAG 634  
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
Qy 635 AGATGATGACTTCTGGCTGCTGAGGAGCTGACAGCTGCTGCTATGGCTTACGCCAAG 694  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
Qy 695 GACTTCAACCGGCTTAAGAAGAAGATGACGCTCCAGAGCTTCCAGTCACCTGCACCTTG 754  
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
Qy 755 GAGGCACATGGTTTC 769  
Db 221 GlyAlaHisGlyPhe 225

RESULT 5  
US-09-106-182-2  
; Sequence 2, Application US/09106182  
; Patent No. 6046035  
; GENERAL INFORMATION:  
; APPLICANT: Shi, Yangu  
; APPLICANT: Ruben, Steve  
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Human Genome Sciences, Inc  
; STREET: 9410 Key West Ave  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: US  
; ZIP: #0850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/106,182  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/051,053  
FILING DATE: 30-JUN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF385  
TELEPHONE: 301-309-8504  
TELEFAX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-106-182-2

Alignment Scores: 3 22e-105 Length: 225  
Pred. No.: 1193.00 Matches: 218  
Score: 98.22% Conservative: 3  
Percent Similarity: 96.89% Mismatches: 4  
Best Local Similarity: 79.64% Indels: 0  
Query Match: 3 Gaps: 0  
DB: 3

US-09-931-704-4 (1-819) x US-09-106-182-2 (1-225)

Qy 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCTATGCACGGTGTGTGG 154  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
Qy 155 CACCTCCCTGCGAGTCCAGCTCTTATCCGACGAGATCCAGCCCTGCCCTCCATC 214  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40  
Qy 215 CAGAAAACCTATGACCTCACCGCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
Qy 275 TACCTGAACCTAGCTGGGGCCCCCTTCAACGAGCGCTGACTTCAATCTCTCGACTGGGG 334  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
Qy 335 CGAGAACTCTCCGAGCGGCGCTCACTTGGAGTGTGGGAGCCCTCAATGACAGG 394  
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
Qy 395 CTGGCGCTGACCCAGAACCTATGAGGGGTACAGTCACCTCTCTGTGTTACTTGGCGCTC 454  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
Qy 455 AACCTGAGCTGCCAGAGCTGAACCTCCGACGCTAGCTGGCGCCCTTCTGTACAGCCTC 514  
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140  
Qy 515 CAGGCGCTGCTGGGAGCATTCAGGTGTCATGGCGGACCTGGCTTACCCACTGCCCCAG 574  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
Qy 575 CCTCTCCAGGAGCTGAGCCAGCTGGGGCCCCCTGGCCCTGCCACAGTACTTCTCCAG 634  
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
Qy 635 AGATGATGACTTCTGGCTGCTGAGGAGCTGACAGCTGCTGCTATGGCTTACGCCAAG 694  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
Qy 695 GACTTCAACCGGCTTAAGAAGAAGATGACGCTCCAGAGCTTCCAGTCACCTGCACCTTG 754  
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
Qy 755 GAGGCACATGGTTTC 769  
Db 221 GlyAlaHisGlyPhe 225



QY 635 AAGATGATGACTTCTGGCTGCTGAGAGAGCTGACAGCTGGCTATGCGCTTACGCAAG 694  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluGlnThrTrpLeuTrpArgSerAlaLys 200  
QY 695 GACTTCAACCGGCTTAAGAGAGATGACGCTCCAGCAGCTTACAGTCCACCTGCGACTTG 754  
Db 201 AspPheAsnArgLeuLysLysMetGlnProAlaAlaValThrLeuHisLeu 220  
QY 755 GAGGCACATGGTTTC 769  
Db 221 GlyAlaHisGlyPhe 225

RESULT 6  
US-08-988-819-2  
; Sequence 2, Application US/08988819  
; Patent No. 6054294  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/988,819  
; FILING DATE: 12-DEC-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442A  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-988-819-2

Alignment Scores:  
Pred. No.: 3,22e-105 Length: 225  
Score: 1193.00 Matches: 218  
Percent Similarity: 98.2% Conservative: 3  
Best Local Similarity: 96.8% Mismatches: 4  
Query Match: 79.6% Indels: 0  
DB: 3 Gaps: 0

US-09-931-704-4 (1-819) x US-08-988-819-2 (1-225)

QY 95 ATGGACCTCCGAGCGGGAGCTGCTGGGGATGTTAGCTTGCCTATGACGCTGCTGG 154  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
QY 155 CACCTCCCTGCGAGTCCAGCTCTTAATCGCAGAGATCCAGGCGCTGGCCCTCCATC 214  
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlySerIle 40  
QY 215 CAGAAAACCTATGACTCACCCGCTACCTGGAGCATCACTCCGAGCTTAGCTGGACC 274  
Db 41 GlnLysThrThrAspLeuThrArgTrpLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 275 TACCTGAACCTACCTGGGGCCCCCTTTCAACAGCCTGACTTCAATCTCTCCTCAGCTGGG 334  
Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80  
QY 335 GCAGAACTCTGCCCCAGGGCCAGGTCACCTTGGAGAGTGTGGCAAGCCTCAATGACAGG 394  
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
QY 395 CTGCGGCTGACCCAGAACCTATGAGGCGTACAGTCACCTCCTGTGTTACTTGCCTGGCCTC 454  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
QY 455 AACCGTCAGGCTGCCACAGCTGAACCTCCAGCTAGCTGGCCCTCTCTGTACACAGCTC 514  
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
QY 515 CAGGGCTCTGCTGGGAGCATTTGACAGTGTATGCGGACGCTTGGCTTACCCACTGCCCTCAG 574  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160  
QY 575 CTCTGCCAGGACTGAGCCAGCTGGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCG 634  
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
QY 635 AAGATGATGACTTCTGGCTGCTGAGAGAGCTGACAGCTGGCTATGCGCTTACGCAAG 694  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
QY 695 GACTTCAACCGGCTTAAGAGAGATGACGCTCCAGCAGCTTACAGTCCACCTGCGACTTG 754  
Db 201 AspPheAsnArgLeuLysLysMetGlnProAlaAlaValThrLeuHisLeu 220  
QY 755 GAGGCACATGGTTTC 769  
Db 221 GlyAlaHisGlyPhe 225

RESULT 7  
US-09-016-534-2  
; Sequence 2, Application US/09016534  
; Patent No. 6143874  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, MING-SHI  
; APPLICANT: ELLIOTT, GARY S.  
; APPLICANT: SARMIENTO, ULLA  
; APPLICANT: SENALDI, GIORGIO  
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: ONE AMGEN CENTER  
; CITY: THOUSAND OAKS  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,534  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/792,019  
; FILING DATE: 03-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOK, ROBERT R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-442B  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-016-534-2

## Alignment Scores:

Pred. No.: 3,22e-105 Length: 225  
Score: 1193.00 Matches: 218  
Percent Similarity: 98.22% Conservative: 3  
Best Local Similarity: 96.89% Mismatches: 4  
Query Match: 79.64% Indels: 0  
DB: 4 Gaps: 0

US-09-931-704-4 (1-819) x US-09-016-534-2 (1-225)

QY 95 ATGGACCTCCGAGGAGGAGCTGCTGGGGATGTTAGCTTGCTATGACGGTGTGTGG 154  
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20  
QY 155 CACCTCCCTGCAGTCCAGCTCTTAATCGACAGGAGATCCAGGCCCTGGCCCTCCATC 214  
Db 21 HisLeuProAlaValProAlaLeuAsnAsgThrGlyAspProGlyProSerIle 40  
QY 215 CAGAAAACCTATGACCTCACCGCTACCTGGAGCATCACTCCGCGAGCTTAGCTGGACC 274  
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60  
QY 275 TACCTGAATACCTGGGGCCCTTTCAAGGACCTGACTTCAATCCTCTCGACTGGGG 334  
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80  
QY 335 CGAGAACTCTGCCAGGCGCAGGTCAACTTGGAGTGTGGGAGCCCTCAATGACAGG 394  
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100  
QY 395 CTGGGCTGACCCAGCAATATGAGCGGTACAGTCACCTCCTGTGTACTTTCGCTGGCCTC 454  
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120  
QY 455 AACCTGACGCTGCCACAGCTGAACTCCGACGTAGCTGGCCCTCTCTGACAGCTC 514  
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140  
QY 515 CAGGGCTGCTGGGAGCATTTGAGGTGTATGCGAGCGCTTGCTACCCACTGCCCCAG 574  
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaLeuGlyTyrProLeuProGln 160  
QY 575 CTTCTCCAGGACTGAGCAGCGCTGGGCGCTGGCCCTGCGCCACAGTACTTCTCCTCAG 634  
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180  
QY 635 AAGATGATGACTTCTGGCTGCTGAGGAGCTGACAGCTGGCTATGCGTTTCAGCCAG 694  
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200  
QY 695 GACTTCAACCGGCTTAAGAAAGATGACAGCTCCAGCAGCTTCAGTCACCTCAGCTTG 754  
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220  
QY 755 GAGGCACATGGTTTC 769  
Db 221 GlyAlaHisGlyPhe 225

## RESULT 8

US-08-642-255-32  
Sequence 32, Application US/08642255  
Patent No. 5773249  
GENERAL INFORMATION:  
APPLICANT: CAPPELLO, Joseph  
APPLICANT: FERRARI, Franco A.  
TITLE OF INVENTION: High Molecular Weight Collagen-Like  
TITLE OF INVENTION: Protein Polymers  
NUMBER OF SEQUENCES: 135

## CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/642,255  
FILING DATE:

## CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: ROWLAND, Bertram I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: A55556-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299 FHT UR  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 330 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-642-255-32

## Alignment Scores:

Pred. No.: 7.32e-06 Length: 330  
Score: 146.00 Matches: 74  
Percent Similarity: 36.36% Conservative: 2  
Best Local Similarity: 35.41% Mismatches: 89  
Query Match: 9.75% Indels: 44  
DB: 1 Gaps: 11

US-09-931-704-4 (1-819) x US-08-642-255-32 (1-330)

QY 25 GCGGCTCGCCCTCCACTCCGCCAGCC---TCTGGAGAGGAGCCGCGCCGCGGCC 81  
Db 35 AlaProGlyProProGlyProProGlyProProGlyProProGlyAlaProGlyProPro 54  
QY 82 GCGCCCGCAG---CCCCATGGACCTCCGAGCAGGCGACTCGTGGGGATCTTAGCTTGCCT 138  
Db 55 GlyProGlyProProGlyProPro-----GlyProAlaGlyProValGlySerPro 72  
QY 139 ATGCAGCGTGTGTGGCACCTCCCTGCAGTGCCA-----GCTCT 177  
Db 73 -----GlyAlaProGlyPro-ProGlyProProGlyProProGlyProProGlyAlaPr 90  
QY 178 TAATCGCAGGAGATCCAGGCCCTGGCCCTCCATCCAGAAACCTATGACCTACCCG 237  
Db 90 oGlyProProGlyProProGlyPro--ProGlyProProGlyProAlaGlyProValGly 109  
QY 238 CTACCTGGAGCATCACTCCGCGAGCTTAGCTGGGACCTACCTGAACTACTGCGGCC 297  
Db 110 SerProGlyAlaProGlyProProGlyProProGlyProProGlyProProGlyAlaPro 129  
QY 298 TTTCAACGAGCCTGACTTCAATCTCTCGACTGGGGGCGAGAACTCTGCCAGGGCCAC 357  
Db 130 -----GlyProProGlyProProGlyProProGlyProProGlyProPro 141  
QY 358 GGTCAACTTGAAGTGTGGCGAAGCTCAATGACAGGCTGGCGCTGACCCAGAACTATGA 417  
Db 142 GlyProAlaGlyProValGlySerPro----- 150  
QY 418 GCGGTACAGTCACTCTCTGTGTACTTGGTGGCCCTCAACCGTCAGGCTGCCACAGCTGA 477



```

; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-475-411A-65

Alignment Scores:
Pred. No.: 7.88e-06 Length: 408
Score: 146.00 Matches: 74
Percent Similarity: 36.36% Conservative: 2
Best Local Similarity: 35.41% Mismatches: 89
Query Match: 9.75% Indels: 44
DB: 4 Gaps: 11

US-09-931-704-4 (1-819) x US-08-475-411A-65 (1-408)
QY 25 GCGGCTCGCCTCCACTCCGAGCC---TCTGGAGAGGAGCGCGCGCGCGCGCC 81
Db 35 AlaProGlyProProGlyProProGlyProProGlyProGlyProGlyProPro 54
QY 82 GCGCCCCAG---CCCCATGGACCTCCGAGCGAGGACTCGTGGGGATGTAGCTTGCCT 138
Db 55 GlyProGlyProProGlyProPro-----GlyProAlaGlyProValGlySerPro 72
QY 139 ATGCACGGTGTGTGGCACCTCCCTGCAGTGCCA-----GCTCT 177
Db 73 -----GlyAlaProGlyPro-ProGlyProProGlyProProGlyProGlyAlaPr 90
QY 178 TAATCGCACAGAGATCGAGCGCTGGCGCTCCATCCAGAAACCTATGACCTCACCG 237
Db 90 oGlyProProGlyProProGlyPro--ProGlyProProGlyProAlaGlyProValGly 109
QY 238 CTACCTGGAGCATCAACTCCGAGCTTAGCTGGACCTACCTGAACCTACTCTGGGGCCCC 297
Db 110 SerProGlyAlaProGlyProProGlyProProGlyProProGlyProProGlyAlaPro 129
QY 298 TTTCACGAGCCTGATTCAATCTCTCGACTGGGGGCGAGAACTCTGCCCGAGGCCAC 357
Db 130 -----GlyProProGlyProProGlyProProGlyProProGlyProPro 141
QY 358 GGTCAACTTGAAGTGTGGCGAAGCTCAATGACAGCGTGGCGCTGACCCAGAACTATGA 417
Db 142 GlyProAlaGlyProValGlySerPro----- 150
QY 418 GCGGTACAGTCACTCTCTGTTACTTGGTGGCTCAACCGTCAGGCTGCCACAGCTGA 477
Db 151 GlyAlaProGlyProProGlyProProGlyProProGlyProProGly----- 166
QY 478 ACTCCGACGTAGCTGGCCCCACTTGTATCAGACCTCCAGGGCTGCTGGCGAGCATTCG 537
Db 167 AlaProGlyProProGlyPro-----ProGlyProProGlyProPro-GlyPro---A 183
QY 538 AGGTGTATGGGACGCTGGCTGACCACTGCCCGCCAGCCTCTCCAGGAGTGA-----CC 594
Db 183 aGlyProValGlySerProGlyAlaProGlyProProGlyProProGlyProProGlyPr 203
QY 595 AGCTGGGCGCTGGCCCT 613
Db 203 oProGlyAlaProGlyPro 209

RESULT 11

US-08-478-029A-65
; Sequence 65, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-029A-65

Alignment Scores:
Pred. No.: 7.88e-06 Length: 408
Score: 146.00 Matches: 74
Percent Similarity: 36.36% Conservative: 2
Best Local Similarity: 35.41% Mismatches: 89
Query Match: 9.75% Indels: 44
DB: 4 Gaps: 11

US-09-931-704-4 (1-819) x US-08-478-029A-65 (1-408)
QY 25 GCGGCTCGCCTCCACTCCGAGCC---TCTGGAGAGGAGCGCGCGCGCGCC 81
Db 35 AlaProGlyProProGlyProProGlyProProGlyProProGlyProProGlyPro 54
QY 82 GCGCCCCAG---CCCCATGGACCTCCGAGCGAGGACTCGTGGGGATGTAGCTTGCCT 138
Db 55 GlyProGlyProProGlyProPro-----GlyProAlaGlyProValGlySerPro 72
QY 139 ATGCACGGTGTGTGGCACCTCCCTGCAGTGCCA-----GCTCT 177
Db 73 -----GlyAlaProGlyPro-ProGlyProProGlyProProGlyProGlyAlaPr 90
QY 178 TAATCGCACAGAGATCGAGCGCTGGCGCTCCATCCAGAAACCTATGACCTCACCG 237
Db 90 oGlyProProGlyProProGlyPro--ProGlyProProGlyProAlaGlyProValGly 109
QY 238 CTACCTGGAGCATCAACTCCGAGCTTAGCTGGACCTACCTGAACCTACTCTGGGGCCCC 297
Db 110 SerProGlyAlaProGlyProProGlyProProGlyProProGlyProProGlyAlaPro 129
QY 298 TTTCACGAGCCTGATTCAATCTCTCGACTGGGGGCGAGAACTCTGCCCGAGGCCAC 357
Db 130 -----GlyProProGlyProProGlyProProGlyProProGlyProPro 141
QY 358 GGTCAACTTGAAGTGTGGCGAAGCTCAATGACAGCGTGGCGCTGACCCAGAACTATGA 417
Db 142 GlyProAlaGlyProValGlySerPro----- 150
QY 418 GCGGTACAGTCACTCTCTGTTACTTGGTGGCTCAACCGTCAGGCTGCCACAGCTGA 477
Db 151 GlyAlaProGlyProProGlyProProGlyProProGlyProProGly----- 166
QY 478 ACTCCGACGTAGCTGGCCCCACTTGTATCAGACCTCCAGGGCTGCTGGCGAGCATTCG 537
Db 167 AlaProGlyProProGlyPro-----ProGlyProProGlyProPro-GlyPro---A 183
QY 538 AGGTGTATGGGACGCTGGCTGACCACTGCCCGCCAGCCTCTCCAGGAGTGA-----CC 594
Db 183 aGlyProValGlySerProGlyAlaProGlyProProGlyProProGlyProProGlyPr 203
QY 595 AGCTGGGCGCTGGCCCT 613
Db 203 oProGlyAlaProGlyPro 209

RESULT 11
```







